A peer-reviewed open-access journal
NeoBiota
Advancing research on alien species and biological invasions

NeoBiota 104: 27–58 (2025)
DOI: 10.3897/neobiota.104.155832

Methods

Deriving inventories of non-native plant species from iNaturalist: Insights from urban centres of the Western Cape, South Africa

Christiaan P. Gildenhuys[1], Luke J. Potgieter[1,2], Cang Hui[3,4,5], David M. Richardson[1]

1 *Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa*
2 *Department of Biological Sciences, University of Toronto, Scarborough, 1265 Military Trail, Toronto, Canada*
3 *Centre for Invasion Biology, Department of Mathematical Sciences, Stellenbosch University, Stellenbosch, South Africa*
4 *Biodiversity Informatics Unit, African Institute for Mathematical Sciences, Cape Town, South Africa*
5 *National Institute for Theoretical and Computational Sciences, Stellenbosch, South Africa*

Corresponding author: Christiaan P. Gildenhuys (christiaan.gildenhuys@gmail.com)

OPEN ACCESS

This article is part of:
Developing lists of alien taxa in the Global South: workflows, protocols, processes, and experiences
Edited by John Wilson, Michele Dechoum, Katelyn Faulkner, Barbara Langdon, Shyama Pagad, Aníbal Pauchard, Hanno Seebens, Tsungai Zengeya, Silvía Ziller

Academic editor: Anibal Pauchard
Received: 14 April 2025
Accepted: 21 October 2025
Published: 13 November 2025

Citation: Gildenhuys CP, Potgieter LJ, Hui C, Richardson DM (2025) Deriving inventories of non-native plant species from iNaturalist: Insights from urban centres of the Western Cape, South Africa. NeoBiota 104: 27–58. https://doi.org/10.3897/neobiota.104.155832

Copyright: © Christiaan P. Gildenhuys et al. This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Abstract

Accurate, up-to-date inventories of non-native species are important to document and improve our understanding of biological invasions globally and inform management decisions. Traditional methods for the collation of inventories are time- and resource intensive, and lists become outdated if not regularly updated. The community science platform iNaturalist can contribute to the collation of regularly updatable ("living") inventories of non-native species. However, robust and transparent workflows are needed to optimise data quality to take full advantage of iNaturalist. We present the semi-Automated Non-Native Inventory Compilation (sANNIC) workflow for the collation and completeness assessment of non-native vascular plant inventories from iNaturalist. The workflow is informed by the World Checklist of Vascular Plants (WCVP) and is used to compare native ranges to a reference area. The utility of the workflow is demonstrated by compiling non-native species inventories of 100 urban centres in the Western Cape province, South Africa. A total of 947 taxa of wild-growing, i.e. casual, naturalised and invasive plants were observed in these urban centres which showed varying levels of sample completeness. Most small towns had too few records for a completeness assessment. Larger urban centres and those near the coast were typically better sampled. This work highlights the potential for iNaturalist to construct non-native species inventories given sufficient coverage and thorough curation.

Key words: Alien plant species, biological invasions, citizen science, community science, inventory, plant invasions, urban ecology

Introduction

Biological invasions are a major threat to biodiversity and human well-being globally and their impacts are projected to increase (Roy et al. 2024). The distribution of non-native species, however, remains poorly understood in many parts of the world (Seebens et al. 2025). Knowledge of which species are present and the non-native species richness of different areas are key indicators for the monitoring and reporting of biological invasions which can allow for informed management (Wilson et al. 2018). Local inventories are important resources to build knowledge

of the distribution of non-native species globally (Van Kleunen et al. 2019), to inform managers and other stakeholders, and to ensure more effective management (Potgieter et al. 2024a). Non-native species inventories are also important for generating and testing hypotheses, answering biogeographical and ecological questions, and building on theories to improve our understanding of biological invasions (Cadotte et al. 2006).

Despite the utility and importance of inventories and species lists, their collation and presentation are often haphazard. A wide range of methods are employed to compile inventories, including systematic surveys, literature searches, and expert opinion (Guézou et al. 2010; Inderjit et al. 2018; Aymerich and Sáez 2019; Nelufule et al. 2023). Differences in methodology complicate meaningful comparisons of inventories and reduce their usefulness in research and management. The estimated completeness of inventories and associated uncertainties are rarely reported. A recent review of non-native species reporting in regional species lists revealed that a large proportion of expert-compiled regional inventories excluded non-native species, and when they were included, they were often not adequately separated from native species (Castro et al. 2023). This review also highlighted that human-dominated areas such as urban centres are under-represented, and that most lists were published as unstructured data (e.g. as a table in a paper) and rarely in machine readable formats or in open databases such as the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/) or other data repositories. Moreover, inventories of non-native species need to be updated regularly to reflect current and emerging invasions, but this is rarely and inconsistently done.

The community science (also known as citizen science) platform iNaturalist (www.inaturalist.org) can help to produce more consistent, easily updatable (or "living") inventories, and allow for the quantification of uncertainty and sampling effort. iNaturalist is an unstructured community science initiative which allows voluntary participants (observers) to add occurrence records of any taxon by uploading one or more images (or sound files) of a single species with the associated time, location, and other information onto the online platform (Mesaglio 2024). These records are later verified by the iNaturalist community (identifiers). The process of species identification is facilitated by an artificial intelligence (AI) image recognition tool which is updated monthly. As of 25 February 2025 this AI tool recognises over 100 000 taxa globally with a reported accuracy rate of close to 90% depending on the taxonomic group and continent (https://www.inaturalist.org/blog/107012-new-computer-vision-model-with-over-100k-taxa; accessed: 10 March 2025). iNaturalist data are uploaded to GBIF monthly, provided the data meet specific licensing criteria (CC0, CC BY, or CC-BY NC) and attain "Research Grade" quality. This quality designation is achievable for wild-growing (i.e. non-cultivated) records that have reached a community consensus, species-level identification and have complete and accurate metadata.

iNaturalist has been widely applied in many parts of the world to study and monitor non-native invasive species. For example, Grattarola et al. (2024) reported on the status of the invasion of *Carpobrotus edulis* in Uruguay using 15 years of iNaturalist records. Potgieter et al. (2024b) demonstrated the utility of iNaturalist in monitoring of urban invasive species by developing a monitoring approach for the invasive polyphagous shot hole borer (*Euwallacea fornicatus*)

in the urban centres of Cape Town and Stellenbosch, South Africa. Young et al. (2021) used iNaturalist to develop automated watch lists for invasive plants in managed areas in the United States. The database of records exported to GBIF has been cited over 6000 times (https://doi.org/10.15468/ab3s5x; accessed 17 March 2025). iNaturalist is increasingly recognised as an indispensable tool for monitoring biodiversity and biological invasions.

Although community science projects provide useful data, limitations can cause misleading results if biases and inaccuracies are not properly addressed (Dobson et al. 2020). Data quality can vary geographically and between taxa (Hochmair et al. 2020; Mesaglio et al. 2023; Alfeus et al. 2024) and can be affected by records of captive/cultivated species if these are not properly flagged as such (Botella et al. 2018; Potgieter et al. 2024a). Much of this variability in data quality can be attributed to variances in identifiers' taxonomic and geographic specialisation and experience in using the platform (Campbell et al. 2023) and to various observer biases (Di Cecco et al. 2021). Thorough data verification is essential for maintaining quality but can be time-consuming.

Workflows that can be regularly and cost-effectively repeated offer the opportunity to streamline the collation of up-to-date inventories, and can standardise and improve reproducibility, transparency, and data quality. For example, Seebens and Kaplan (2022) developed a workflow to Downscale Alien Species Checklists using Occurrence records (DASCO). They integrate the GloNAF database (Global Naturalized Alien Flora; Van Kleunen et al. 2019) with occurrence records from GBIF and the Ocean Biodiversity Information System to create local-scale inventories. However, inventories produced through this and other similar workflows (Young et al. 2021; Saffer et al. 2024) rely on predefined checklists or inventories (such as GloNAF); they therefore have the same limitations related to the collation of inventories as discussed above. Several recent studies have also highlighted shortcomings of databases like GloNAF (Gildenhuys et al. 2024; Potgieter et al. 2024a). The quality of the workflow output also depends on the quality of the input occurrence records (e.g. from GBIF), and many workflows fail to include independent verification of record accuracy. One way of improving quality is to employ truncation, i.e. removing species with observations below a certain threshold. However, this approach discards valuable data for assessing inventory completeness that requires accurate estimates on the numbers of singletons and doubletons (the number of species only observed once or twice in the assemblage respectively) (Chao et al. 2020) and potentially also discards data for taxa at the early stage of invasion which are critical for early detection and monitoring. Alternative workflows that address such limitations by incorporating careful checking of records are therefore required.

Given the inconsistent sampling effort of opportunistic data, reporting on the likely completeness of inventories generated from iNaturalist is vital. Tools for assessing sample completeness and diversity estimation of incomplete samples could address this need (Hsieh et al. 2016; Chao et al. 2020). These tools have already been applied in several studies using opportunistic data from iNaturalist and other community science platforms (Callaghan et al. 2022a; Gorta et al. 2023; Shen et al. 2023; Li et al. 2024; Richardson and Potgieter 2024). However, care must be taken to apply such techniques appropriately, acknowledging the opportunistic nature of the data collection. Nonetheless, some measure of completeness needs to be incorporated into workflows.

Urban centres provide a good focus for the development of such a workflow. Although urban centres are not the primary focus for many iNaturalist users (Thompson et al. 2023), many records are nonetheless collected in urban habitats, near where most observers live (Beninde et al. 2023; Dimson et al. 2023). Focus on invasions in urban centres has increased in the last decade (Salomon Cavin and Kull 2017). Many studies have highlighted the facilitative role of human-aided pathways, cultivation, and novel habitats associated with urban centres in the introduction and naturalisation of non-native species (Donaldson et al. 2014; Faulkner et al. 2016; Van Kleunen et al. 2018; Potgieter and Cadotte 2020; Potgieter et al. 2020, 2024c; Palit et al. 2024). In South Africa, iNaturalist has experienced a large uptake, with over 2.7 million records of plants as of 12 October 2024, and most (1.7 million) of these records are concentrated in the Western Cape province, a disproportionate number of which originate from urban centres (https://www.inaturalist.org/projects/vascular-plants-of-western-cape-urban-areas; accessed 3 March 2025). To date, protocols for effective management of invasive species in urban centres in South Africa are lacking and more information on the distribution of non-native species across urban centres is needed to guide management efforts (Potgieter et al. 2020). Western Cape urban centres therefore make for a good case study of the proposed workflow.

We aim to develop a semi-automated workflow, the "semi-Automated Non-Native Inventory Compilation" (sANNIC), for deriving non-native plant inventories and assessing their completeness from iNaturalist data. To this end we: a) develop an automated R-based script "Automated Non-Native Inventory Compilation" (ANNIC) to automatically build preliminary non-native species lists; b) integrate these lists into iNaturalist projects to thoroughly curate each species and ensure the highest possible data quality and optimal use of iNaturalist; and c) propose a protocol for the evaluation of species list completeness using principles of Chao completeness assessment. We then apply this workflow in 100 urban centres in the Western Cape province, South Africa, to produce an openly available and constantly updating living inventory of non-native species in urban centres. Finally, we discuss the implications, limitations and transferability of the proposed workflow.

Methods

Here we describe the proposed semi-Automated Non-Native Inventory Compilation (sANNIC) workflow (Fig. 1). It consists of four steps: 1) acquiring data from iNaturalist; 2) compiling a non-native inventory from the data informed by the World Checklist of Vascular Plants (WCVP) using the ANNIC R script; 3) creating and curating an iNaturalist project of the non-native inventory; and 4) conducting a completeness assessment of the inventory produced. The sequence and execution of steps are informed by decision points (diamonds in Fig. 1). For example, if the data are deemed of sufficient quality after the second step, the workflow proceeds to the fourth step; if not, it proceeds to the third step and then back to the first and so forth. The goal of the sANNIC workflow is to facilitate the production of non-native plant inventories and encourage manual curation of the data in an iterative manner. We also discuss possible alternative approaches at each step.

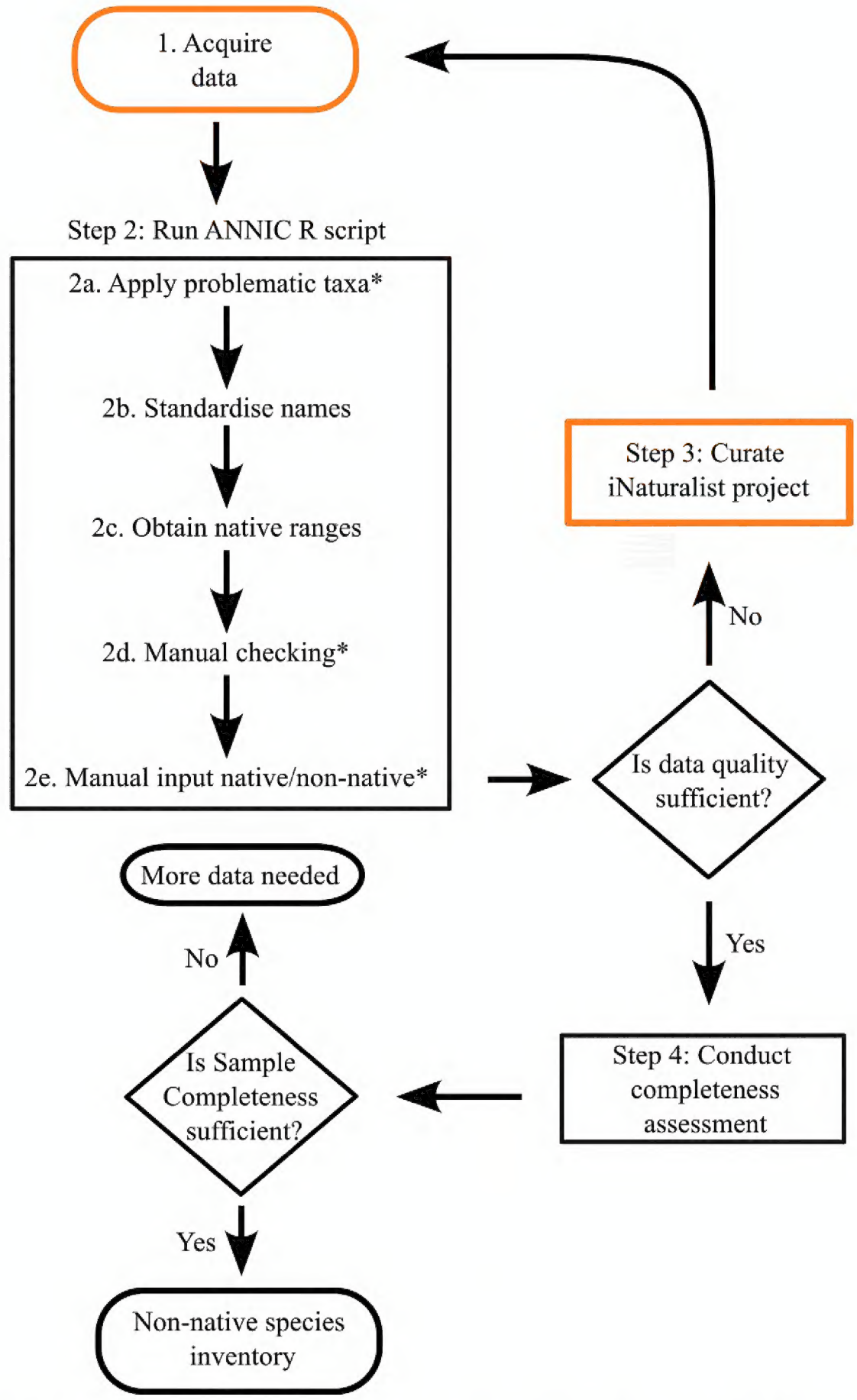


Figure 1. Representation of the semi-Automated Non-Native Inventory Compilation (sANNIC) workflow for the derivation of inventories of non-native plant taxa from iNaturalist. Steps outlined in a black square represent those conducted in R. Steps highlighted in light orange represent those steps that require working in iNaturalist. Diamond shapes represent decision points. Rounded shapes represent the outcome of the workflow. Asterisks (*) indicate optional steps.

Step 1: Acquire data

The area and taxon of interest first need to be defined on iNaturalist. The simplest option is to create a project to capture the relevant information. Areas of interest can be uploaded (if not already present) as 'places' which can be used to refine a search query or used as a geographic filter in a project. The taxonomic group of interest can be specified using a search query in the project settings (see Mesaglio 2024 for more details). 'Places' can be uploaded onto iNaturalist in a KML file format, however, certain requirements must first be met: the user creating the place must have made more than 50 observations on iNaturalist; the place cannot contain more than 100 000 observations before creation; the place area cannot exceed

~ 700 000 km^2; the file size cannot exceed 1MB; and no more than three places can be added by a single user in a day (Mesaglio 2024). As many of these requirements can be quite limiting, we recommend using existing places where possible.

There are several options for exporting data from iNaturalist depending on the amount of data needed. Data can be accessed using the export function on the iNaturalist website which allows for up to 200 000 records per query. Various Application Programming Interfaces (APIs) are available, for example R packages such as 'rinat', which can export up to 10 000 observations per day (Barve et al. 2022; https://api.inaturalist.org/v1/docs/). Many places, including all 'standard places', in iNaturalist are 'checklist' enabled, which allows for the export of a list of all Research Grade taxa recorded within the place boundaries (Mesaglio 2024). This is useful if the number of records of interest far exceeds the iNaturalist export limits. Alternatively, data can be sourced from second parties such as GBIF, however, GBIF only exports Research Grade records and those with particular licencing, representing only a subset of the total number of records on iNaturalist. The data are also only sent to GBIF once a month.

Although our workflow is tailored to work with observations exported directly through the iNaturalist website, it can be modified to function with different data sources. For our workflow the following columns need to be included in the iNaturalist export: id (a unique identifier for each observation), observedon (the standardised date the observation was made), user_id (a unique identifier for each user), quality_grade (including "Research Grade" and "needs id"), captive_cultivated (whether the observation is flagged as captive or cultivated), longitude and latitude (the coordinate location associated with each observation), taxon_id (a unique taxonomic identifier), taxon_family_name (family name of the taxon as recorded on iNaturalist), taxon_subfamily_name (subfamily name of the taxon as recorded on iNaturalist), taxon_genus_name (genus name of the taxon), taxon_species_name (species name of the taxon if identified to species level), taxon_hybrid_name (hybrid name of the taxon if identified as a hybrid), taxon_subspecies_name (subspecies name of the taxon if identified as a subspecies). Additional columns like the url (a link to the web address of each observation), positional_accuracy (precision of the coordinates) and others are not required but can be useful to include in any export. Any additional columns can be added to the input as needed.

Step 2: Automated Non-Native Inventory Compilation (ANNIC)

The Automated Non-Native Inventory Compilation (ANNIC) R script automatically classifies species as native or non-native based on the World Checklist of Vascular Plants (WCVP) by using iNaturalist occurrence records as input and assisting users in manually reviewing species that cannot be assigned automatically. Written in R version 4.4.3 (R Core Team 2025), it is available as an R script on Zenodo (https://zenodo.org/records/15210704) and the latest version is available on GitHub (https://github.com/christiaan-g/sANNIC_workflow). It requires the 'tidyverse' packages to conduct data manipulation (Wickham et al. 2019), 'rWCVPdata' to access the latest snapshot of the WCVP (Govaerts 2024), and 'rWCVP' to provide several functionalities to deal with the WCVP data (Brown et al. 2023).

2a: Problematic taxa (optional)

Some taxa, regardless of image quality of the record, cannot be consistently identified to species level from iNaturalist records. This is mainly because of unclear distinctions between species within a genus or other taxonomic group, difficulty in distinguishing similar species, absence of sufficiently high-resolution images showing key traits required for species-level identification, hybridisation, difficulty in identifying taxa at the juvenile stage, and a lack of available taxonomic knowledge to identify records to species level. Such taxa are considered "problematic taxa"; the R script allows for these taxa to be collapsed to a higher taxonomic rank, including genus, family, subfamily, or any manual specification. These taxa will be included if present in the dataset regardless of their quality grade; special attention is therefore needed during the curation process (see section 2d. "Manual checking"). This step is optional but when working with iNaturalist data we recommend assessing the presence of problematic taxa. All problematic taxa in this workflow are assumed to be non-native.

2b: Standardise names (matching with the WCVP)

The World Checklist of Vascular Plants (WCVP) is a continuously updated database of described vascular plant species curated by the Royal Botanic Gardens, Kew (Govaerts et al. 2021). The database represents a global consensus of vascular plant taxonomy and contains information on species authorship, distribution, lifeform, and climate descriptions. As our proposed workflow focuses on vascular plants, we propose using this database to standardise nomenclature and access native ranges (see section 2c. "Obtaining native ranges"). We recommend collapsing subspecies into species as many subspecies are not well documented in the WCVP. If some subspecies are deemed necessary to include in a given study these can be specified as problematic taxa (see section 2a. "Problematic taxa").

Names are matched to the WCVP to ensure congruity between the different data sources, using exact and fuzzy matching algorithms in the 'rWCVP' package (Brown et al. 2023). The matching algorithms return multiple matches for taxa that have multiple entries in the WCVP. In this case, matches are resolved algorithmically if they are the only accepted taxon or accepted synonym among multiple matches. The matching algorithms also return poor matches, namely phonetic fuzzy matches using the "metaphone" algorithm (Howard 2020), edit distance fuzzy matches using the highest Levenshtein edit distance (Levenshtein 1966), and no matches (see Brown et al. 2023 for more details). Phonetic matches are retained if their edit similarity exceeds 0.9. All other poor matches, taxa for which no match was kept, or taxa for which multiple matches were kept, are flagged for subsequent manual checking. All poor matches are visualised in Fig. 2a, b.

2c: Obtain native ranges

The WCVP stores distribution data according to the World Geographic Scheme for Recording Plant Distributions (WGSRPD; Brummitt 2001). The finest scale available through the WCVP is the 'botanical country' (tdwg3 in Darwin Core) which usually aligns with political boundaries of countries or provinces/states. The choice of region (or regions) of interest is flexible and can be specified in the workflow using the tdwg3

or 'botanical country' code. These codes can be found using the "get_wgsrpd3_codes" function in the 'rWCVP' package. Taxa recorded in the WCVP as native within the botanical country (or countries) of interest are classified as "native" and taxa recorded as having native ranges entirely outside this area or which have introduced ranges in the region(s) of interest are classified as "nonnative". If a taxon is classified as both native and introduced across multiple selected regions of interest, the native status takes precedence. Some taxa listed in the WCVP have no native or introduced range specifications; in these cases taxa will be flagged for manual checking (Fig. 2a, b).

2d: Manual checking (optional)

All taxa previously flagged for manual checking because of unresolved matching with the WCVP or missing native range information in the WCVP, can be manually assessed. While this step is optional, omitting it will result in the exclusion of certain taxa (Fig. 2a, b). The workflow generates a spreadsheet of taxa requiring manual verification, which can be edited using any spreadsheet software. The user is prompted to check whether each suggested match to the WCVP should be accepted for those taxa which have poor matches and tag species as native or non-native. This can be done by filling in three columns: 1) "match_correct" which requires a Boolean TRUE if the suggested match to the WCVP is correct or FALSE if it is not correct, 2) "nativeness" which requires the user to specify either "native" or "nonnative" for each taxon, and 3) "checked" which requires the user to enter "checked" if the relevant taxon has been reviewed and should be reincorporated into the workflow, otherwise the taxon will be removed. Upon completion, the spreadsheet will be read back into R and integrated with the remaining data.

2e: Apply manual native and non-native input (optional)

Some species listed in the WCVP have erroneous distribution data and can therefore be falsely flagged as native or non-native (Fig. 2e, f). To account for this, the R script allows the user to manually input taxa which, if present, will be tagged as native or non-native (as specified by the user) regardless of their distributions as shown in the WCVP. These incorrect classifications can be identified during the curation process in Step 3, or by using an *a priori* list of known native or non-native taxa. Although optional, this step is highly recommended to minimise any errors that might arise in the course of the workflow.

Step 3: Load and curate the data on an iNaturalist project

After a non-native species list is produced, a new project can be created on iNaturalist and set to collate all taxa in the list. This is done in the project settings by manually entering each species name. Due to limitations on the number of taxa that can be added to an iNaturalist project, some taxa can be listed at genus level where all species in that genus are non-native to the reference area. For example, in South Africa, we included the genera *Acacia*, *Eucalyptus*, and *Pinus* for which all species are non-native. If the number of taxa is still too large, one solution is to create multiple co-occurring projects each targeting a number of species in the allowed limit. For example, these can be grouped by alphabetical names of families.

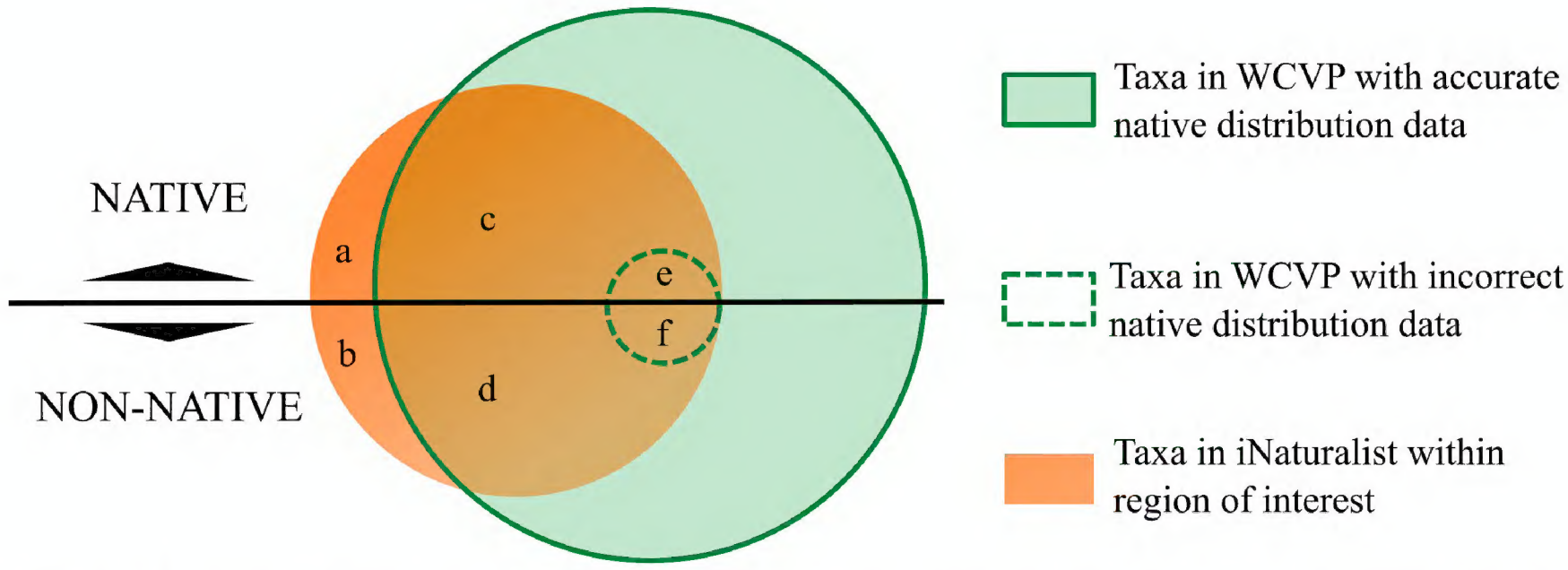


Figure 2. Venn diagram visualising the output of the ANNIC-R script. Proportions are not to scale. The large green circle represents taxa listed in the World Checklist of Vascular Plants (WCVP) and for which native distribution information is available. The orange circle represents taxa listed in our region of interest in iNaturalist. Species above the black line are native, those below the line are non-native. The small dashed green circle represents species which were listed in the WCVP with incorrect native distribution information. **a , b.** Represent native and non-native species for which good matches were not found in the WCVP or for which distribution information is not available; **c, b.** Represent species listed in our project for which good matches and distribution information was available in the WCVP; **e, f.** Represent taxa for which the distribution information in the WCVP was incorrect.

Once an iNaturalist project has been created, records need to be systematically reviewed to ensure, and improve, data quality. This should be done within the iNaturalist platform because improving the records will ultimately contribute to improving the overall data quality on the platform. We also recommend engaging with users on the platform to aid in the collection and identification of relevant records. Many users on iNaturalist may also be relevant stakeholders in the region of interest whose activities, livelihoods, or well-being could be directly or indirectly affected by biological invasions. The workload of this manual curation step varies depending on the region and the taxon of interest. Some taxa are already well curated on iNaturalist (Mesaglio et al. 2023). To determine whether certain taxa are well curated, the proportion of records identified to Research Grade level is a good general guide (Mesaglio et al. 2023). Records can be identified as accurately as possible by communicating with taxonomic experts through iNaturalist and using species descriptions and keys in both peer-reviewed and grey literature where available. See similar steps described by Richardson and Potgieter (2024).

We recommend investing effort in evaluating whether iNaturalist records represent captive/cultivated records. This workflow is intended for wild-growing plants, but iNaturalist records are often not correctly tagged as captive/cultivated. For plants, records can be tagged as cultivated based on considerations including, signs of cultivation or care (including trimming, weeding, support, planting pattern, etc), life stage (e.g., seedlings are more likely to be wild-growing/self-sown than large adults which may have been planted long ago), and environmental setting (e.g. plants in a garden are more likely to be cultivated than plants in a riparian zone or semi-natural vegetation). Other information such as prior knowledge of the species, location, observer notes, and communication with observers can be useful in separating wild-growing from cultivated plants. For discussion of considerations, see Gildenhuys et al. (2024), Potgieter et al. (2024a), and Richardson and Potgieter (2024). When cultivation status is unclear and cannot be ascertained for an observation, it is generally advised to assume it is wild-growing, following a precautionary principle. Special care should be taken to ensure that rare species,

i.e. taxa with only one or two records are correctly identified and appropriately tagged as captive/cultivated or wild-growing. The frequency of these taxa will inform completeness assessments and can underestimate completeness if the number of rare species is inflated (Shen et al. 2023).

Step 4: Completeness assessment

Sample coverage is a concept first developed by Alan Turing (1912–1954) and later published by Good (1953). It describes the proportion of all individuals in an assemblage represented by the species captured in the sample. It was later modified and improved by Chao and Jost (2012) and applied to ecological contexts. It is calculated using the sample size, and singletons and doubletons of a sample, i.e. number of species only observed once or twice in the assemblage respectively. Chao et al. (2020) later expanded the concept of sample coverage, incorporating the concept of Hill numbers to create a continuous sample completeness profile of order q. Sample completeness for q = 0 is simply the proportion of species that have been observed (conventional meaning of sample completeness); for q = 1, it becomes sample coverage (the proportion of the total number of individuals for these detected species in the entire assemblage); for q = 2, it weighs a detected species by its squared detection probability when computing the proportion, thus disproportionally favouring more common species.

We suggest calculating sample completeness measures, using either abundance or incidence-based methods, at orders q = 0,1,2 using the *Completeness* function in the 'iNext.4steps' package in R (Chao et al. 2020). Sample completeness 95% confidence intervals are calculated via bootstrapping using the *Completeness* function. The resulting completeness profile can be visually assessed to evaluate Sample Completeness. Our interpretation here is based on the interpretation provided by (Chao et al. 2020). The values of all completeness measures typically fall between 0 and 1, where 0 is a perfectly incomplete sample and 1 a perfectly complete sample (with completeness defined in terms of q). When the profile falls (i.e. declining sample completeness with increasing orders of q), sampling is considered inadequate for a completeness assessment, and we have no confidence in the sample completeness measures. A well sampled area will have a rising slope with small confidence intervals and sample completeness measure of orders q = 1,2 close to 1. A complete sample will have a flat slope with sample completeness measure of all orders of q close to 1. Very few samples are likely to achieve this extent of completeness. Sample completeness of order q = 1 can be used as a general sample completeness measure when the slope is rising.

The major assumption of completeness estimation is the independence of samples, which is not always the case with opportunistically collected data. Sometimes observers will record the same species multiple times in quick succession, ignoring all other species, thereby violating the assumption of independence. This can severely bias sample completeness estimation, especially when there are few observations. We recommend data coarsening when this is the case so that observations by the same observers, of the same species, in the same area on the same day are assumed to be duplicates and then removing the additional records. This effectively produces a more conservative sample coverage estimate but accounts for cases where extreme biases are likely. Other stricter or more lenient corrections can be made depending on the needs of the user. Note that this will not remove the

general bias generated from opportunistic data but will at least mediate some of the more severe cases. If there is inflation in the number of singletons and doubletons compared to a 'true' sample or if the heterogeneity of observation probability of unseen species is highly skewed compared to the 'true' heterogeneity, the sample completeness can still be underestimated or overestimated. Nonetheless, the proposed completeness assessment can be a highly useful tool to assess how complete the iNaturalist data are with appropriate interpretation.

Western Cape case study

We showcased the application of the sANNIC workflow to derive non-native species inventories from iNaturalist by applying it to 100 urban centres in the Western Cape, South Africa. We created and uploaded several urban shapefiles (Fig. 3) in KML file format to iNaturalist to delineate the urban centres of the Western Cape (see Appendix 1 for details on the creation of the shapefiles). We created an iNaturalist project collecting records from all 100 urban centres using the uploaded shapefiles and refined the taxonomic group of interest to vascular plants (available at: https://www.inaturalist.org/projects/vascular-plants-of-western-cape-urban-areas).

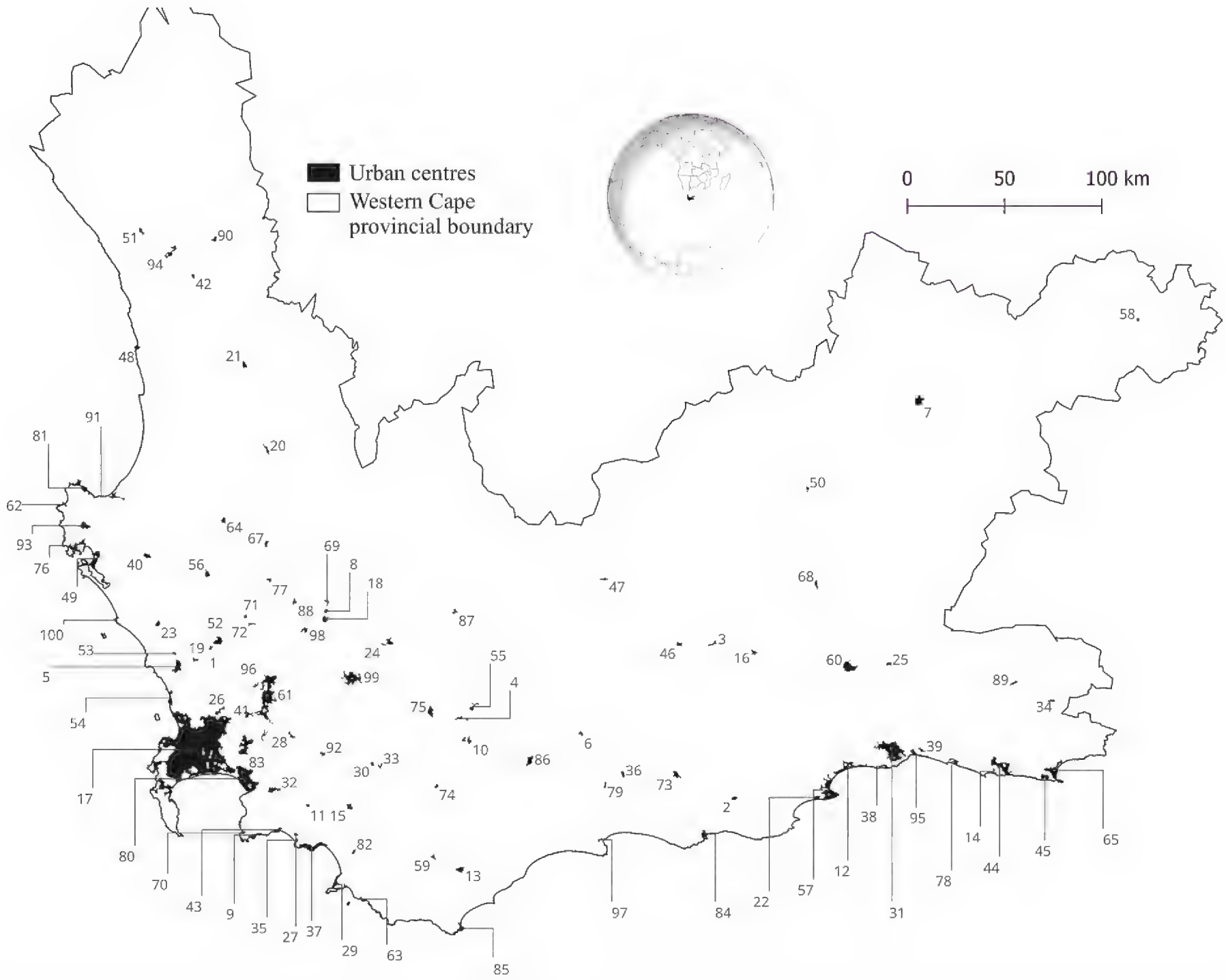


Figure 3. Urban centres of the Western Cape, South Africa, included in the study to develop a workflow to produce inventories of non-native plant taxa. Boundaries of urban centres were based on the built-up area from the South African National Land Cover database (available at: https://egis.environment.gov.za/sa_national_land_cover_datasets). Numbers refer to the urban centres listed in Table 2. See Appendix 1 for more details.

In this project, we are interested in all non-native species which have wild-growing (i.e. non-cultivated) records. This includes all species in category C1 or higher according to the Blackburn et al. (2011) framework. This includes all escaping, casual, naturalising, and invasive species (Richardson et al. 2000). A finer scale differentiation from a source like iNaturalist is difficult as key variables such as abundance, spread, and whether a population is self-sustainable cannot easily be ascertained from opportunistically collected presence-only data. Though the number of records of a particular species in a particular area can be a useful indicator of whether the population is advanced along the naturalisation-invasion continuum, this is not necessarily the case and may be severely biased. Differentiating between cultivated and wild-growing records presents a more manageable challenge. Individual records can be tagged as likely wild-growing or cultivated using secondary data in the images uploaded and other metadata for each record (see Step 3 for details).

We exported over 300 000 records of vascular plants in the project directly from the iNaturalist website recorded after 2017, with at least one photo, and with an accuracy unknown or less than 1000 m using a series of queries on the export page. Most records (>95%) contributed to iNaturalist in the Western Cape urban centres were observed after 2017 and any records before this time may not reflect the current non-native flora. Observations with an accuracy of greater than 1000 m were excluded as there is a high likelihood that these records fall outside the urban delineations. Approximately 2% of records with accuracy values exceeded 1000 m. We assume, thus, that records with no accuracy recorded fell within the urban delineations, seeing this as an acceptable omission we included all records with no accuracy recorded. We included all the variables listed in Step 1 for export and added the following variables to allow for easier quality checking only: updated_at (the datetime that changes were last made to the observation), license (license or waiver chosen by observer for the observation), url (a link to the web address of each observation), num_identification_agreements (number of identifications that agree with the observer's identification), num_identification_disagreements (number of identifications that disagree with the observer's identification), positional_accuracy (precision of the coordinates), and scientific_name (the scientific name for each taxon).

All subspecies were collapsed into species, and we considered several problematic taxa which could not be easily and reliably identified and were collapsed to the specified taxonomic rank in Appendix 2. Genera collapsed were: *Alocasia* (unclear distinction between various species), *Amaranthus* (unclear species distinctions), *Avena* (species not distinguishable from typical iNaturalist images), *Brugmansia* (Difficult to distinguish species from typical iNaturalist images), *Fraxinus* (unclear species distinctions), *Gamochaeta* (unclear species distinctions), *Vitis* (unclear species distinctions and not distinguishable from typical image quality), *Salsola* (unclear species distinctions), *Sambucus* (species difficult to distinguish from typical iNaturalist images), *Narcissus* (hybridisation), *Neltuma* (species difficult to distinguish from typical iNaturalist images), *Hedera* (unclear species distinction), *Pinus* (difficult to identify at juvenile stage), *Eucalyptus* (difficult to identify at juvenile stage or without flowers/fruits), and *Sagina* (unclear species distinctions). The subfamily Bambusoideae (Poaceae) was collapsed due to species not being distinguishable from typical image quality. The Section Oenothera of the genus *Oenothera* was collapsed as species level identifications within this section are extremely difficult and many species regularly hybridise.

Note that our list of "problematic taxa" could change in the future as more information becomes available about these taxa or as we become aware of other difficult taxa. As we could not attain Research Grade quality for these records due to the requirement for species-level identification, each record was carefully checked to ensure it was correctly identified to at least the rank to which it was aggregated.

All taxon names were matched with the latest snapshot of the WCVP (updated May 2024) using the 'rWCVP' package. Any taxa which achieved poor matches were manually checked to see whether the suggested matches were correct. Most of these taxa with poor matches were either taxa not yet listed in the WCVP, obscure or artificial hybrids, typographic differences (e.g. "*Opuntia monacantha*" vs. "*O. monacanthos*") and non-standard names (e.g. "*Hermania* sp12"). We checked the native/non-native status of all taxa that needed manual checking by conducting a search for the species name on Google Scholar or referring to local authorities such as the South African National Biodiversity Institute (SANBI) website (https://pza.sanbi.org/; last accessed 25 March 2025). Hybrids were included as non-native if at least one of the parent taxa were non-native to our study region (e.g., *Rubus* × *fultus* which is a hybrid of the native *R. pinnatus* and non-native *R. affinis*).

The WGSRPD splits South Africa into four botanical countries: Cape Provinces (CPP; comprising the modern provinces of the Eastern, Northern, and Western Cape), Free State (OFS), KwaZulu-Natal (NAT), and Northern Provinces (TVL; comprising the modern provinces of Gauteng, Limpopo, Mpumalanga, and Northwest). For the purposes of this case study, CPP was used as the region of interest and only taxa recorded in the WCVP as native to the CPP were considered native. Taxa not recorded in the CPP or recorded as introduced in the CPP were considered non-native. We manually uploaded all taxon names compiled in this non-native list to a project in iNaturalist titled "Non-native plants of Western Cape urban areas" (https://www.inaturalist.org/projects/non-native-plants-of-western-cape-urban-areas; last accessed 3 March 2025). We manually curated the project by assessing the cultivation status and correct identification of records as discussed in Step 3. Any native species which were found to be incorrectly included in the project were removed and were recorded as manual native species for future iterations of the workflow (see section 2e). Any non-native species which were incorrectly excluded from the project (based on our knowledge of local invaders) were added and recorded as manual non-native species. The completeness assessment was conducted for the Western Cape urban centres in its entirety and for each of the urban centres which had adequate records (Appendix 3).

In our Western Cape case study, we aimed to identify the most prevalent taxa within our region of interest. As a proxy of prevalence, we ranked each taxon according to the number of observations in an urban centre because species that are more prevalent are also likely to be observed more frequently on iNaturalist. This consideration alone, however, biases the list in favour of the best-sampled urban centres. One approach to address this is to count the number of urban centres in which each taxon was found; however, this does not account for the prevalence of species within each urban centre. To account for sampling variations among urban centres, we thus derived a prevalence index that is calculated as the sum of a taxon's inverse rank squared for each urban centre weighed by the urban centre's sample coverage (Sample Completeness, q = 1). Let PIj be the prevalence index of species j; n the number of urban centres; SCi the sample coverage at urban centre i (i.e. sample completeness

of order q = 1; Appendix 3); and Rankij the rank of species j in urban centre i, with Rankij = ∞ if species j does not occur in urban centre i. The prevalence index reads,

$$PI_j = \sum_{i=1}^{n} \frac{SC_i}{\sqrt{\text{Rank}_{ij}}}$$

We reported some of the patterns observed in the non-native flora across the Western Cape urban centres by determining the most represented families and the most common life form descriptions extracted from the WCVP. These were determined by number of species, number of total observations, and number of species weighted by their prevalence index scores. We also compared the taxon list to the GloNAF list of known naturalised or invasive species for both the region and globally (version 2.01 available at: 10.5281/zenodo.13235357). We collapsed all listed problematic species (Appendix 2) in GloNAF in the same fashion as for the inventory to make for a fair comparison.

Results

Applying our workflow to the Western Cape urban centres yielded a total of 4467 vascular plant taxa, excluding the collapsed problematic taxa. Matches in the WCVP were found for 4405 taxa, of which 4364 were exact matches, 34 were edit distance fuzzy matches and seven were phonetic matches. Multiple matches were found for 496 taxa, and no matches were found for the remaining 62 taxa. Most multiple matches could be resolved algorithmically. All taxa with edit distance fuzzy matches, unresolved multiple matches, no matches, and for which no distribution information was available on the WCVP, were flagged for manual checking. A total of 126 matches of 108 taxa had to be manually checked (Fig. 1: Step 2d).

Taxa erroneously flagged in the WCVP as non-native in our region of interest were removed from the list, namely *Pteridium aquilinum*, *Tarchonanthus littoralis*, *Tetragonia fruticosa*, *Scabiosa columbaria*, *Solanum rubetorum*, *Senecio pinifolius*, and *Salicornia pilansii*. We also added non-native taxa which were erroneously excluded from the inventory, namely *Cotula australis* and *Rumex acetosella*. All these taxa had erroneous native range information in the WCVP, highlighting the importance of this reviewing step.

We consulted with many local and international taxonomic experts through iNaturalist. Many identifications were made by the authors DMR, CPG, and LJP, reflecting our own expertise and personal consultations and reference to the taxonomic literature. Remaining uncertainties in species identifications, for various reasons, were accounted for by including them in our list of problematic taxa (Appendix 2). The list reflects our own uncertainties in species' identification at the time of writing and may change as our understanding of plant identification improves.

After curation of the data, we improved the percentage of Research Grade records in the project (excluding problematic taxa) from 86.4% to 94%. Given the variation in image quality on iNaturalist, we expect that it will be difficult to increase this number further. The nomenclature between iNaturalist and the WCVP are largely congruent as iNaturalist adopts the WCVP nomenclature to a large extent (Mesaglio 2024). Most differences between WCVP and iNaturalist taxonomical nomenclature were synonyms, obscure hybrids, typographic differences, e.g. "*Polygala garcini*" vs "*P. garcinii*", and non-standard names, e.g. "*Hermania* sp12". Taxonomic names reported in this paper follow the iNaturalist nomenclature.

A total of 947 wild-growing non-native taxa with 61 259 corresponding records were observed in iNaturalist. The completeness profile of the 100 urban centres, as a whole, indicates a high sample completeness across all measures. The rare species sensitive sample completeness q = 0, i.e. the proportion of estimated species which were observed, was 0.7823 (95% CI: 0.7314–0.8331); Sample coverage, q = 1, was 0.9958 (95% CI: 0.9953–0.9962); and the abundant species sensitive measure, q = 2, was close to 1. This indicates that almost all relatively abundant species were recorded and as many as 78% of all non-native species, including very rare species, were recorded. A total of 219 taxa were only observed once and 91 observed only twice. The top 100 species ranked by the Prevalence Index are shown in Table 1.

The most represented families by number of species were Fabaceae, Asteraceae, Poaceae, Solanaceae, and Myrtaceae (Suppl. material 1: figs S1–S3). When comparing families by cumulative prevalence index or number of records, it had relatively little effect on the most represented families. The most common life form description by number of species was annual, followed by tree, shrub or tree, and

Table 1. Top 100 most prevalent non-native plant taxa in urban centres of the Western Cape, South Africa derived from iNaturalist using a semi-automated workflow. Taxa marked with an asterisk (*) represent mainly invaders of patches of natural or semi-natural vegetation at the edge or embedded within the boundaries of urban centres.

Taxon	Prevalence index	Number of urban centres found	Total number of observations
*Acacia cyclops**	11.665	37	722
*Acacia saligna**	11.465	45	1794
Plantago lanceolata	10.271	58	1584
Echium plantagineum	9.052	41	1551
*Acacia mearnsii**	8.362	33	617
Lantana camara	7.263	29	583
Tropaeolum majus	6.234	29	1139
*Pinus**	6.200	30	943
Malva arborea	5.755	25	256
Lagurus ovatus	5.431	21	590
Briza maxima	5.280	28	755
Hypochaeris radicata	5.125	29	1015
Solanum mauritianum	5.084	22	472
Ricinus communis	4.897	35	435
Schinus terebinthifolia	4.836	29	646
Myoporum insulare	4.792	28	186
Trifolium repens	4.726	25	495
Physalis peruviana	4.560	25	380
Cenchrus clandestinus	4.548	29	791
Oenothera sect. Oenothera	4.517	25	339
Silene gallica	4.465	24	333
Sonchus oleraceus	4.358	33	743
Nephrolepis cordifolia	4.349	23	362
Melia azedarach	4.292	36	543
Medicago polymorpha	4.207	31	427
Spartium junceum	3.884	24	400

Taxon	Prevalence index	Number of urban centres found	Total number of observations
*Gaudium laevigatum**	3.860	19	235
*Eucalyptus**	3.851	23	699
Erodium moschatum	3.751	25	500
Oenothera lindheimeri	3.743	28	237
Catharanthus roseus	3.731	24	363
Cenchrus setaceus	3.717	22	467
Cirsium vulgare	3.703	26	381
Lysimachia loeflingii	3.695	29	262
Cortaderia selloana	3.621	17	384
Anredera cordifolia	3.560	21	264
Trifolium angustifolium	3.426	18	498
Stenotaphrum secundatum	3.424	20	426
Helminthotheca echioides	3.416	25	489
*Opuntia ficus-indica**	3.315	26	170
*Acacia melanoxylon**	3.297	18	346
Verbena bonariensis	3.256	22	322
Avena	3.154	26	497
Hedera	2.966	19	668
Arundo donax	2.959	27	321
Lotus subbiflorus	2.940	21	269
Solanum nigrum	2.882	27	378
Fumaria muralis	2.839	23	436
Coleus barbatus	2.799	22	426
Thunbergia alata	2.789	19	389
Euphorbia peplus	2.763	20	381
Limonium sinuatum	2.742	17	56
Taraxacum officinale	2.727	22	438
Cyrtomium falcatum	2.671	16	123
Melilotus indicus	2.659	23	201
Lysimachia arvensis	2.629	20	145
Oxalis corniculata	2.602	21	407
Chenopodiastrum murale	2.592	23	222
Datura stramonium	2.587	22	197
Atriplex semibaccata	2.516	23	81
Melaleuca viminalis	2.414	17	445
Myoporum montanum	2.409	23	133
Paspalum urvillei	2.377	19	248
Tecoma stans	2.367	19	78
Brassica tournefortii	2.306	20	187
Euphorbia terracina	2.292	15	334
Opuntia monacantha	2.288	16	126
Erigeron karvinskianus	2.267	16	282
Phytolacca octandra	2.258	20	164
Bidens pilosa	2.231	16	173
Hibiscus trionum	2.196	16	135

Taxon	Prevalence index	Number of urban centres found	Total number of observations
Vinca major	2.182	20	351
Pittosporum undulatum	2.177	17	192
Cestrum laevigatum	2.176	17	246
Ipomoea indica	2.065	18	411
Acacia longifolia	2.055	15	464
Canna × hybrida	2.040	20	153
Verbascum virgatum	2.024	20	172
Erigeron bonariensis	2.016	19	412
Vicia benghalensis	1.984	15	202
*Hakea sericea**	1.956	16	62
Paraserianthes lophantha	1.928	15	406
Raphanus raphanistrum	1.925	19	374
Cerastium glomeratum	1.920	19	172
Rumex crispus	1.907	18	191
Foeniculum vulgare	1.899	19	264
Limonium perezii	1.799	14	145
Kalanchoe delagoensis	1.795	15	50
Vicia sativa	1.792	18	182
Sesbania punicea	1.788	20	282
Myoporum laetum	1.786	18	103
Erigeron sumatrensis	1.784	17	250
Nothoscordum gracile	1.769	17	269
Amaranthus	1.766	21	278
Acacia podalyriifolia	1.756	16	87
Geranium molle	1.744	17	201
Portulaca oleracea	1.692	17	241
Nasturtium officinale	1.679	10	194
Solanum lycopersicum	1.652	15	121
Diplotaxis muralis	1.636	14	67

perennial (Suppl. material 1: figs S4–S6). Of the 947 taxa that were recorded only 404 (43%) were represented as naturalised or invasive in GloNAF in the Western Cape, 588 (62%) were represented in GloNAF in South Africa, and a further 267 (28%) were recorded as naturalised or invasive elsewhere in the world. Only 91 taxa recorded were not represented in GloNAF at all (Suppl. material 1: fig S7).

The distribution of records across the urban centres was highly uneven, with Cape Town having the highest number of records (31 034), followed by Stellenbosch (13 689), Somerset West (3 156), George (2 721), and Mossel Bay (1191). The observed species richness was also highest in Cape Town (686), followed by Stellenbosch (410), Somerset West (383), George (376), and Hermanus (253). According to sample coverage q = 1, the best sampled urban centres were Cape Town (0.994), followed by Stellenbosch (0.991), Somerset West (0.963), George (0.949), and Mossel Bay (0.941). Completeness estimates, number of records and species richness for all urban centres are shown in Table 2. See Appendix 3 for the completeness profile for all urban centres for which completeness assessment was possible. Of the 100 urban centres included in this study, 85 had two

or more records (the theoretical minimum number of records required for species estimation). Fifty-seven towns had 20 or more records and only 25 had more than 100 records (Table 2). Thirteen towns had a sample coverage (q = 1) estimate of greater than 0.8, and 37 towns had an abundant species sensitive sample completeness (q = 2) greater than 0.8. This indicates that though most towns (87%) were not sampled sufficiently to have a high degree of completeness, many towns did have a large proportion of highly abundant species sampled (Fig. 4). These species are likely to correspond to the most invasive species in the region.

Table 2. One hundred urban centres in the Western Cape, South Africa, included in the study to derive an inventory of non-native plant taxa from iNaturalist using a semi-automated workflow. SC.LCL is the lower confidence limit (95%) and SC.UCL is the upper confidence limit of the sample coverage (sample completeness for q = 1).

Nr	Urban centre	Area (km²)	Number records	Number species	Sample Coverage	SC.LCL	SC.UCL
1	Abbotsdale	2.32	0	0	NA	NA	NA
2	Albertinia	3.48	3	3	NA	NA	NA
3	Amalienstein & Zoar	3.31	1	1	NA	NA	NA
4	Ashton	4.93	7	5	NA	NA	NA
5	Atlantis	16.44	20	15	NA	NA	NA
6	Barrydale	2.91	37	29	0.281	0.139	0.423
7	Beaufort West	14.56	45	18	NA	NA	NA
8	Bella Vista	2.55	0	0	NA	NA	NA
9	Betty's Bay	8.48	753	161	0.900	0.881	0.918
10	Bonnievale	7.24	0	0	NA	NA	NA
11	Bot River	1.74	47	30	0.374	0.175	0.573
12	Brakrivier	24.96	781	195	0.868	0.845	0.891
13	Bredasdorp	7.50	46	36	0.323	0.154	0.492
14	Brenton	2.51	175	84	0.743	0.688	0.797
15	Caledon	6.23	55	31	0.569	0.456	0.682
16	Calitzdorp	3.86	25	17	NA	NA	NA
17	Cape Town	846.78	31034	686	0.994	0.994	0.995
18	Ceres	6.65	8	8	NA	NA	NA
19	Chatsworth	2.95	0	0	NA	NA	NA
20	Citrusdal	4.20	0	0	NA	NA	NA
21	Clanwilliam	4.79	14	7	0.601	0.353	0.850
22	Dana Bay	4.84	213	81	0.746	0.703	0.788
23	Darling	4.16	37	19	0.731	0.585	0.878
24	De Doorns	9.05	0	0	NA	NA	NA
25	Dysseldorp	3.01	1	1	NA	NA	NA
26	Fisantekraal	6.39	2	1	NA	NA	NA
27	Fisherhaven	1.94	564	75	0.926	0.908	0.945
28	Franschhoek	4.28	35	25	0.479	0.338	0.621
29	Gansbaai	12.14	106	53	0.684	0.594	0.774
30	Genadendal	2.35	10	8	NA	NA	NA
31	George	74.79	2721	376	0.949	0.941	0.957
32	Grabouw	9.92	73	53	0.425	0.298	0.553
33	Greyton	2.67	175	102	0.581	0.502	0.659
34	Haarlem	2.58	0	0	NA	NA	NA
35	Hawston	2.41	30	15	NA	NA	NA
36	Heidelberg	3.80	7	6	NA	NA	NA

Nr	Urban centre	Area (km²)	Number records	Number species	Sample Coverage	SC.LCL	SC.UCL
37	Hermanus	23.32	1110	253	0.904	0.888	0.919
38	Herolds Bay	3.63	54	33	0.461	0.355	0.568
39	Hoekwil	3.07	97	59	0.529	0.409	0.650
40	Hopefield	5.86	10	8	NA	NA	NA
41	Klapmuts	6.81	11	10	NA	NA	NA
42	Klawer	2.54	4	4	NA	NA	NA
43	Kleinmond	5.16	763	215	0.837	0.819	0.855
44	Knysna	35.30	575	202	0.818	0.787	0.849
45	Kranshoek	8.62	88	43	0.635	0.519	0.750
46	Ladismith	3.98	18	16	NA	NA	NA
47	Laingsburg	2.77	2	2	NA	NA	NA
48	Lambert's Bay	3.52	0	0	NA	NA	NA
49	Langebaan	19.68	54	39	NA	NA	NA
50	Leeu Gamka	1.98	3	2	NA	NA	NA
51	Lutzville	4.19	1	1	NA	NA	NA
52	Malmesbury	12.19	30	25	0.246	0.017	0.475
53	Mamre	2.15	6	6	NA	NA	NA
54	Melkbosstrand	8.17	68	33	0.661	0.537	0.785
55	Montagu	6.34	35	27	0.358	0.167	0.550
56	Moorreesburg	6.08	0	0	NA	NA	NA
57	Mossel Bay	39.68	1192	182	0.941	0.930	0.952
58	Murraysburg	2.05	1	1	NA	NA	NA
59	Napier	2.92	111	48	0.749	0.666	0.833
60	Outshoorn	23.63	64	37	NA	NA	NA
61	Paarl & Wellington	96.54	451	172	0.789	0.754	0.823
62	Paternoster	2.15	51	30	0.558	0.426	0.689
63	Pearly Beach	2.35	133	29	0.595	0.448	0.742
64	Piketberg	4.67	19	9	0.563	0.313	0.812
65	Plettenberg Bay	24.59	627	136	0.840	0.812	0.868
66	Pniel & Kylemore	4.01	151	60	0.660	0.581	0.739
67	Porterville	3.70	16	9	0.293	0.000	0.692
68	Prince Albert	4.33	19	13	0.457	0.180	0.734
69	Prince Alfred Hamlet	2.93	3	3	NA	NA	NA
70	Pringle Bay	2.93	127	62	0.673	0.591	0.734
71	Riebeek West	2.00	42	31	0.406	0.230	0.583
72	Riebeek-Kasteel	2.45	67	39	0.606	0.454	0.758
73	Riversdale	7.63	61	39	0.564	0.436	0.692
74	Riviersonderend	2.21	14	12	NA	NA	NA
75	Robertson	10.04	7	7	NA	NA	NA
76	Saldanha Bay	16.47	12	9	NA	NA	NA
77	Saron	2.52	0	0	NA	NA	NA
78	Sedgefield	6.66	153	73	0.718	0.656	0.779
79	Slangrivier	2.31	1	1	NA	NA	NA
80	Somerset West	78.96	3156	383	0.963	0.958	0.969
81	St Helena Bay	13.46	20	15	NA	NA	NA
82	Stanford	2.76	63	49	NA	NA	NA
83	Stellenbosch	33.57	13689	410	0.991	0.990	0.993
84	Still Bay	8.77	244	110	0.776	0.730	0.824
85	Struisbaai & L'Agulhas	7.23	28	16	0.689	0.526	0.852
86	Swellendam	10.48	165	86	0.689	0.631	0.747

Nr	Urban centre	Area (km²)	Number records	Number species	Sample Coverage	SC.LCL	SC.UCL
87	Touws River	3.22	3	3	NA	NA	NA
88	Tulbagh	3.72	20	18	NA	NA	NA
89	Uniondale	3.09	24	16	0.463	0.210	0.716
90	Vanrhynsdorp	3.69	17	13	0.426	0.143	0.709
91	Velddrif	8.56	37	25	0.448	0.226	0.671
92	Villiersdorp	3.25	5	5	NA	NA	NA
93	Vredenburg	11.44	11	9	NA	NA	NA
94	Vredendal	10.85	6	6	NA	NA	NA
95	Wilderness	8.16	332	110	0.804	0.766	0.842
96	Windmeul	2.98	16	13	NA	NA	NA
97	Witsand	2.22	43	22	0.675	0.545	0.804
98	Wolseley	4.86	14	10	0.521	0.223	0.819
99	Worcester	34.95	45	32	0.422	0.262	0.582
100	Yzerfontein	3.46	46	29	0.527	0.379	0.674

Discussion

Main strengths of the workflow

Our study confirms the utility of iNaturalist data for compiling species inventories. The inventory generated from the workflow ultimately contributes to the iNaturalist platform by encouraging species identification and data curation which improves the data for future use (e.g., for monitoring and detection). The project derived from the workflow (https://www.inaturalist.org/projects/non-native-plants-of-western-cape-urban-areas) is freely accessible. As additional observations are posted in our region of interest, the project will automatically be updated. All data and R code used in this paper are freely available on Zenodo (https://zenodo.org/records/15210704) and GitHub (https://github.com/christiaan-g/sANNIC_workflow). The sample completeness estimation flags both areas that are adequately sampled and those that are under-sampled and can guide the allocation of resources for adaptive biodiversity monitoring (Henrys et al. 2024). The Prevalence Index provides a unique way of ranking species based on their local abundance and broader extent within the study area considering sample completeness.

Our workflow further quantifies some of the uncertainty inherent in the compilation of non-native species inventories. We used a sample completeness profile to assess the inventory completeness of urban centres (Fig. 4) which is based on the biodiversity sampling measures recommended by Chao et al. (2020). When the completeness profile increases and the measures of order 1 and 2 are similar and high, it suggests that the area is well-sampled. Alternatively, if there is high uncertainty in the different measures or if the profile decreases, it can indicate insufficient or unrepresentative sampling. Species diversity can be compared between areas which have a similar sample coverage. This can be done using sample rarefaction and extrapolation or asymptotic estimation (Hsieh et al. 2016).

Although our data were sourced from iNaturalist and our code is designed for this platform, the workflow can easily be adapted to incorporate data from other sources. For example, data can be sourced from other popular community science platforms such as Pl@ntNet (https://plantnet.org/en/) and eBird (https://ebird.org/home) or from traditional databases of herbarium and museum specimens. We recommend using the proposed workflow alongside other sources of

data to improve the comprehensiveness of the inventory. The data from many of these sources are already available on GBIF. Each data source will have their own data quality requirements which should be carefully evaluated and used to guide appropriate adaptations of the workflow. We used the WCVP to extract native range information for each species because it is the most extensive database available for vascular plants. Other sources of information (e.g. IUCN native range shapefiles) could be used for other taxonomic groups such as birds and mammals. However, for many taxonomic groups, no compiled databases are available, and knowledge remains limited.

The proposed sANNIC workflow functions fundamentally differently from workflows such as DASCO (Seebens and Kaplan 2022). While other workflows often attempt to downscale or synthesise already existing inventories focussed on species that are already advanced on the naturalisation-invasion continuum (Richardson et al. 2000), sANNIC is aimed at building an inventory from scratch using iNaturalist data, focussing on species at all stages along the naturalisation-invasion continuum. As a result, it generates a more comprehensive inventory and is less likely to overlook non-native species recorded on iNaturalist compared to other workflows. A limitation to the workflow is that it requires intensive curation and knowledge of the taxa of interest and can be time consuming. Our workflow can potentially be used together with other workflows and applications such as *infinitylists* which allow users to rapidly generate species checklists for a particular area from GBIF data (Mesaglio et al. 2025). We stress the need for data curation to ensure the robustness of generated inventories.

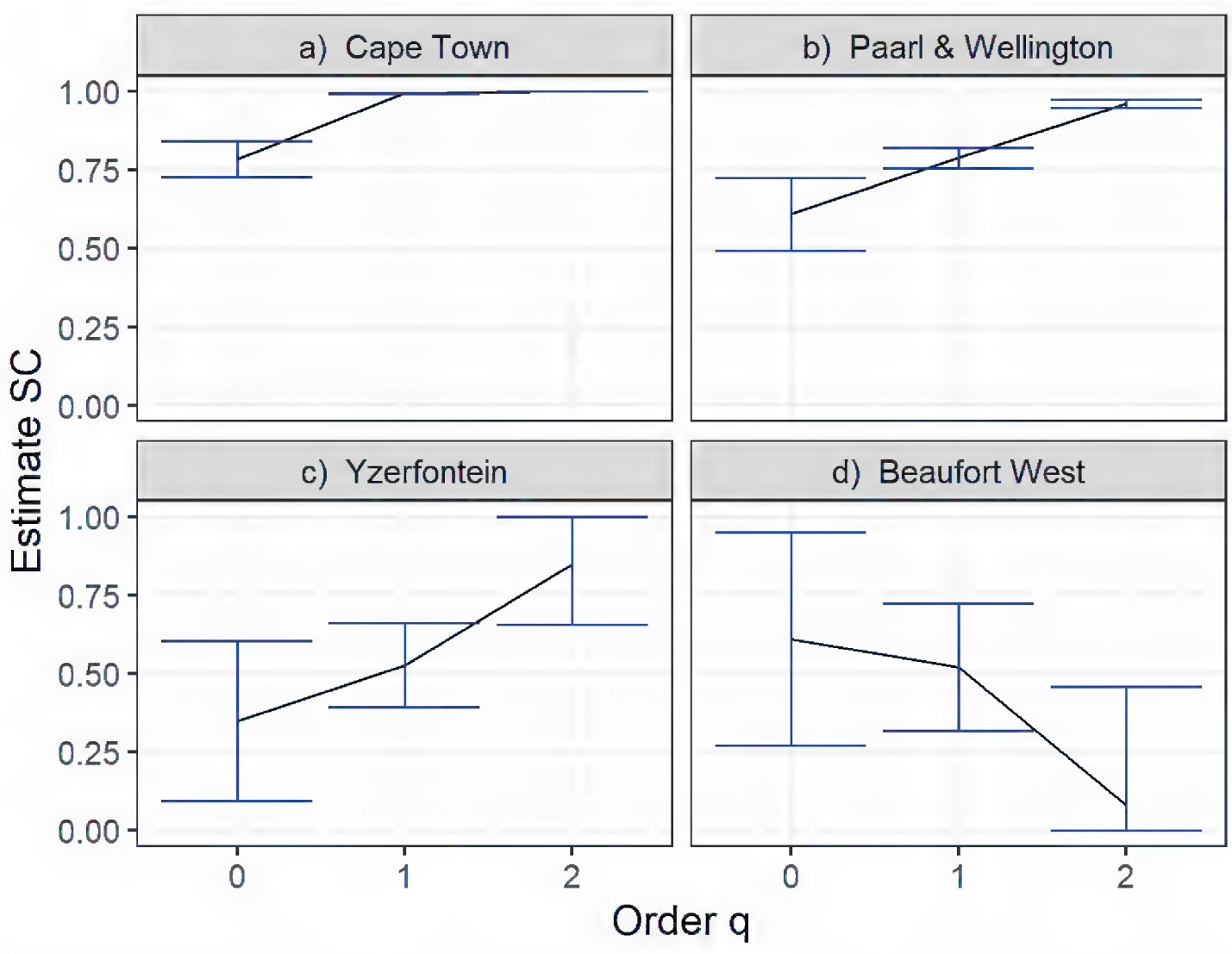


Figure 4. Completeness profiles for four urban centres in the Western Cape ranging from most complete (top left) to least complete (bottom right). Panels (**a–d**) represent urban centres with the following profiles. **a.** Highly complete sample coverage (q = 1, > 0.8); **b.** Lower completeness, but with a high proportion of highly abundant species sampled (q = 2, > 0.8); **c.** Poor sampling, but a rising completeness profile indicates that sample completeness can at least be estimated; **d.** Poor sampling and a falling completeness profile indicates that estimates of sample completeness may be unreliable.

Limitations and practical recommendations

The level of curation required for the application of the sANNIC workflow in different regions is likely to vary, given the differences in identification and observation efforts across regions (Campbell et al. 2023). The drawbacks of using iNaturalist include dealing with difficult taxonomic groups for which many species cannot be identified given the typical image quality on iNaturalist (Ackland et al. 2024); potential crypticity and hybridisation with native species pose additional challenges. Moreover, the workflow does not differentiate the status of species along the naturalisation-invasion continuum but can be used as a foundation for further refinement. The prevalence index can be used as a proxy of advancement along the naturalisation-invasion continuum, as abundant and widespread species are more likely to be invasive than rarer "casual" species. However, this may not necessarily be the case, and the introduction status, e.g. categorisation according to the Blackburn et al. (2011) framework, needs to be determined separately from the workflow.

We observed a highly uneven sampling effort across the urban centres of the Western Cape as most towns had insufficient records for a completeness assessment. Sampling effort is skewed toward larger urban centres closer to the coast, though there were some exceptions. This spatial skewness is typical for iNaturalist and other unstructured community science platforms as users are more likely to reside in these areas or visit them for recreational purposes. The skewness provides a challenge for species monitoring across the area of interest, as many areas remain under sampled. We accounted for the uneven sampling by using the prevalence index which considers the rank of a species at each urban centre where it was sampled, weighed by the Sample Coverage of that centre. Using this approach, we produced a list of the most prevalent species in urban centres of the Western Cape while minimising potential biases introduced by spatial heterogeneity. The prevalence index can also be adapted for use in a grid- or observer-based framework to account for spatial heterogeneity or potential observer bias. However, this index may still be biased by general observer preferences and/or the inherent recognisability of species (Mesaglio et al. 2023).

The number of records in some areas could be increased by City Nature Challenges (Di Cecco et al. 2021); these events challenge users to collect as many records as possible over several days. Although records should be collected consistently throughout the year, seasonal heterogeneity can be mitigated by assessing whether observations are distributed year-round, promoting data collection during off-seasons, or applying data coarsening where sufficient data are available. Furthermore, the capacity for boosting research grade records, increasing the number of high-quality records, is limited by the number and quality/experience of identifiers. We recommend engaging with existing identifiers on the platform and experts not yet engaged with iNaturalist to overcome the shortage of identifiers (Callaghan et al. 2022b).

Although estimating sample completeness is essential for interpreting the data, several limitations need to be considered. Uneven spatial or seasonal sampling within urban centres can affect sample coverage and influence how accurately the number of species can be estimated. For example, if observations are strongly concentrated during a particular time of year, the sample coverage will be overestimated since only species visible during that time of year will be observed. Strong spatial autocorrelation of records may result in an overestimation of sample coverage and underestimation of species richness if species occurrence is highly heterogeneous. We encourage caution when using sample completeness measures, especially in areas with few records and

strongly encourage the reporting of uncertainty of sample completeness measures using bootstrapping, as is available in the 'iNEXT' packages (Chao et al. 2020). These tools, however, have their limitations and comparing diversity of areas with differing levels of sample completeness is less robust and therefore not recommended. Even when sample completeness is low for a given area, the resulting species list remains valid and can be valuable for some ecological questions or management purposes.

Western Cape inventory case study

The list of species generated in this case study has many potential uses and implications for the regional understanding of plant invasions in urban areas of the Western Cape of South Africa. Many taxa in our list are among the most common urban plant invaders globally (Li et al. 2024) but there are many exceptions. Many species in our list are largely invaders of natural or semi-natural vegetation embedded within or adjoining the urban areas, e.g. *Acacia* and *Pinus* species (Richardson et al. 2020). Invaders of natural habitats may benefit strongly from the movement of propagules (intentional or unintentional) through urban centres. Other species owe their presence to widespread plantings, mainly as ornamentals, and subsequently self-seed into urban green spaces (Donaldson et al. 2014; Gildenhuys et al. 2024; Louw et al. 2024; Milton and Dean 2025). Urban centres are often donors of non-native species to surrounding areas (McLean et al. 2018). Novel urban habitats and disturbances created by urbanisation may aid in the establishment of many species within urban centres. Potgieter et al. (2024c) highlight the different ways in which urban habitats can shape biological invasions.

The best represented families, (Fabaceae, Asteraceae, and Poaceae) in our case study are not surprising; these are generally the best represented plant families in non-native floras globally (Suppl. material 1: figs S1–S3). However, there were a few families such as Myrtaceae which are disproportionally represented in the non-native flora of this region. This may partly be because of the long history of introduction of many species in this family to South Africa (Poynton 2009). Plant life forms were varied (Suppl. material 1: figs S4–S6) with the most common life form being annual and/or perennial herbs followed by shrubs and/or trees. The high representation of herbaceous life forms is to be expected in highly modified urban environments where a ruderal strategy may be more advantageous (Frazee et al. 2019). Further work is needed to elucidate the patterns and drivers of invasion in the study area. Of the 947 taxa, 856 (90%) were represented in GloNAF globally, but only 404 (43%) were listed as being naturalised in the Western Cape in GloNAF. Another 185 (20%) were listed as naturalised in South Africa in GloNAF, but not in the Western Cape, and 91 taxa were not represented in GloNAF at all (Suppl. material 1: fig S7). This highlights some of the gaps in the GloNAF database especially at the sub-national level. The absence of some of the taxa in GloNAF but present in our inventory may be attributable to differences in categorising species as "casual" rather than "naturalised" or "invasive" (sensu Richardson et al. 2000), but others have simply been overlooked in previous regional assessments.

Future directions

We believe our workflow provides the foundation for further research into the non-native species of any region with sufficient iNaturalist coverage. If an inventory with a high coverage for a particular area is available, research can focus on clarifying the introduction status of each species, introduction pathways, possible impacts,

and patterns of establishment and invasion within the region of interest. We highlight the necessity of thorough curation of the records directly on iNaturalist and believe that our workflow could be used alongside other sources of data such as systematic surveys to produce high quality inventories of non-native species.

Acknowledgements

This research was supported by the Centre for Invasion Biology and the Department of Botany and Zoology, Stellenbosch University. We are deeply grateful to everyone who contributed observations and identifications to the iNaturalist project. These include but are not limited to Clare Archer (@sedgesrock), Brian du Preez (@mr_fab), Sandra Falanga (@sandraf), Bianke Fouche (@biankefouche), Jeremy Gilmore (@jeremygilmore), David Hoare (@dhoare), Suzaan Kritzinger-Klopper (@zaniekk), Caroline Mashau (@carolinemashau), Phil McLean (@fynbosphil), Erick Munro (@erickmunro), Tony Rebelo (@tonyrebelo), Norbert Sauberer (@brothernorbert), Charles Stirton (@charles_stirton), Diana Studer (@dianastuder), Hedi Stummer (@venturefoth), Shaun Swanepoel (@shauns), Pieter Winter (@pieterwinter), Mark van Dalsen (@vandalsen), @linkie, and many more. A full list of identifiers and observers that contributed to the project can be found at https://www.inaturalist.org/projects/non-native-plants-of-western-cape-urban-areas.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Use of AI

No use of AI was reported.

Funding

CPG is supported by a postgraduate scholarship from the National Research Foundation (PMDS230601112319). LJP acknowledges support from the Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, and from the Natural Sciences and Engineering Research Council of Canada (grant 386151). CH is supported by the NRF (grant 89967) and the European Union's Horizon Europe Research and Innovation Programme (B3 – Biodiversity Building Blocks for policy, ID 101059592). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the EU nor the EC can be held responsible for them.

Author contributions

Data cleaning was performed by DMR, CPG, and LJP. Supervision was performed by DMR, CH, and LJP. Writing of associated software was performed by CPG. All authors contributed to the conceptualisation and writing of the manuscript.

Author ORCIDs

Christiaan P. Gildenhuys https://orcid.org/0009-0006-5226-3250

Luke J. Potgieter https://orcid.org/0000-0001-7790-2721
Cang Hui https://orcid.org/0000-0002-3660-8160
David M. Richardson https://orcid.org/0000-0001-9574-8297

Data availability

All code and data used in this study are available in Zenodo (10.5281/zenodo.15210704) and the latest versions will be made available in GitHub (https://github.com/christiaan-g/sANNIC_workflow). All additional figures referred to are available in the Suppl. material 1.

References

Ackland SJ, Richardson DM, Robinson TB (2024) A method for conveying confidence in iNaturalist observations: A case study using non-native marine species. Ecology and Evolution 14: e70376. https://doi.org/10.1002/ece3.70376

Alfeus M, Irish J, Birkhofer K (2024) Recognition and completeness metrics from iNaturalist and GBIF can inform future citizen science and research projects: A case study on arthropods in Namibia. Biodiversity and Conservation. https://doi.org/10.1007/s10531-024-02981-z

Aymerich P, Sáez L (2019) Checklist of the vascular alien flora of Catalonia (northeastern Iberian Peninsula, Spain). Mediterranean Botany 40: 215–242. https://doi.org/10.5209/mbot.63608

Barve V, Hart E, Guillou S (2022) rinat: Access "iNaturalist" Data Through APIs. https://cran.r-project.org/web/packages/rinat/index.html [November 4, 2024]

Beninde J, Delaney TW, Gonzalez G, Shaffer HB (2023) Harnessing iNaturalist to quantify hotspots of urban biodiversity: The Los Angeles case study. Frontiers in Ecology and Evolution 11: 983371. https://doi.org/10.3389/fevo.2023.983371

Blackburn TM, Pyšek P, Bacher S, Carlton JT, Duncan RP, Jarošík V, Wilson JRU, Richardson DM (2011) A proposed unified framework for biological invasions. Trends in Ecology & Evolution 26: 333–339. https://doi.org/10.1016/j.tree.2011.03.023

Botella C, Joly A, Bonnet P, Monestiez P, Munoz F (2018) Species distribution modeling based on the automated identification of citizen observations. Applications in Plant Sciences 6: e1029. https://doi.org/10.1002/aps3.1029

Brown MJM, Walker BE, Black N, Govaerts RHA, Ondo I, Turner R, Nic Lughadha E (2023) rWCVP: A companion R package for the World Checklist of Vascular Plants. The New Phytologist 240: 1355–1365. https://doi.org/10.1111/nph.18919

Brummitt RK (2001) World geographical scheme for recording plant distributions. ed. 2. Hunt Inst. for Botanical Documentation, Pittsburgh, 1–137.

Cadotte MW, Murray BR, Lovett-Doust J (2006) Ecological Patterns and Biological Invasions: Using Regional Species Inventories in Macroecology. Biological Invasions 8: 809–821. https://doi.org/10.1007/s10530-005-3839-4

Callaghan CT, Bowler DE, Blowes SA, Chase JM, Lyons MB, Pereira HM (2022a) Quantifying effort needed to estimate species diversity from citizen science data. Ecosphere 13: e3966. https://doi.org/10.1002/ecs2.3966

Callaghan CT, Mesaglio T, Ascher JS, Brooks TM, Cabras AA, Chandler M, Cornwell WK, Ríos-Málaver IC, Dankowicz E, Dhiya'ulhaq NU, Fuller RA, Galindo-Leal C, Grattarola F, Hewitt S, Higgins L, Hitchcock C, Hung K-LJ, Iwane T, Kahumbu P, Kendrick R, Kieschnick SR, Kunz G, Lee CC, Lin C-T, Loarie S, Medina MN, McGrouther MA, Miles L, Modi S, Nowak K, Oktaviani R, Olewe BMW, Pagé J, Petrovan S, Saari C, Seltzer CE, Seregin AP, Sullivan JJ, Sumanapala AP, Takoukam A, Widness J, Willmott K, Wüster W, Young AN (2022b) The benefits of contributing to the citizen science platform iNaturalist as an identifier. PLOS Biology 20: e3001843. https://doi.org/10.1371/journal.pbio.3001843

Campbell CJ, Barve V, Belitz MW, Doby JR, White E, Seltzer C, Di Cecco G, Hurlbert AH, Guralnick R (2023) Identifying the identifiers: How iNaturalist facilitates collaborative, research-relevant data generation and why it matters for biodiversity science. Bioscience 73: 533–541. https://doi.org/10.1093/biosci/biad051

Castro A, Ribeiro J, Reino L, Capinha C (2023) Who is reporting non-native species and how? A cross-expert assessment of practices and drivers of non-native biodiversity reporting in species regional listing. Ecology and Evolution 13: e10148. https://doi.org/10.1002/ece3.10148

Chao A, Jost L (2012) Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. Ecology 93: 2533–2547. https://doi.org/10.1890/11-1952.1

Chao A, Kubota Y, Zelený D, Chiu CH, Li CF, Kusumoto B, Yasuhara M, Thorn S, Wei CL, Costello MJ, Colwell RK (2020) Quantifying sample completeness and comparing diversities among assemblages. Ecological Research 35: 292–314. https://doi.org/10.1111/1440-1703.12102

Di Cecco GJ, Barve V, Belitz MW, Stucky BJ, Guralnick RP, Hurlbert AH (2021) Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. Bioscience 71: 1179–1188. https://doi.org/10.1093/biosci/biab093

Dimson M, Berio Fortini L, Tingley MW, Gillespie TW (2023) Citizen science can complement professional invasive plant surveys and improve estimates of suitable habitat. Diversity & Distributions 29: 1141–1156. https://doi.org/10.1111/ddi.13749

Dobson ADM, Milner-Gulland EJ, Aebischer NJ, Beale CM, Brozovic R, Coals P, Critchlow R, Dancer A, Greve M, Hinsley A, Ibbett H, Johnston A, Kuiper T, Le Comber S, Mahood SP, Moore JF, Nilsen EB, Pocock MJO, Quinn A, Travers H, Wilfred P, Wright J, Keane A (2020) Making Messy Data Work for Conservation. One Earth 2: 455–465. https://doi.org/10.1016/j.oneear.2020.04.012

Donaldson JE, Hui C, Richardson DM, Robertson MP, Webber BL, Wilson JRU (2014) Invasion trajectory of alien trees: The role of introduction pathway and planting history. Global Change Biology 20: 1527–1537. https://doi.org/10.1111/gcb.12486

Faulkner KT, Robertson MP, Rouget M, Wilson JRU (2016) Understanding and managing the introduction pathways of alien taxa: South Africa as a case study. Biological Invasions 18: 73–87. https://doi.org/10.1007/s10530-015-0990-4

Frazee LJ, Aronson MFJ, Kattge J, Struwe L (2019) Hardscape floristics: Functional and phylogenetic diversity of parking-lot plants. Applied Vegetation Science 22: 573–581. https://doi.org/10.1111/avsc.12450

Gildenhuys CP, Potgieter LJ, Richardson DM (2024) The genus *Quercus* (Fagaceae) in South Africa: Introduction history, current status, and invasion ecology. South African Journal of Botany 167: 150–165. https://doi.org/10.1016/j.sajb.2024.01.066

Good IJ (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40: 237–264. https://doi.org/10.1093/biomet/40.3-4.237

Gorta SBZ, Callaghan CT, Samonte F, Ooi MKJ, Mesaglio T, Laffan SW, Cornwell WK (2023) Multi-taxon biodiversity responses to the 2019–2020 Australian megafires. Global Change Biology 29: 6727–6740. https://doi.org/10.1111/gcb.16955

Govaerts R [Ed.] (2024) WCVP: World Checklist of Vascular Plants. Facilitated by the Royal Botanic Gardens, Kew. https://doi.org/10.34885/nswv-8994

Govaerts R, Nic Lughadha E, Black N, Turner R, Paton A (2021) The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. Scientific Data 8: 215. https://doi.org/10.1038/s41597-021-00997-6

Grattarola F, Rodríguez-Tricot L, Zarucki M, Laufer G (2024) Status of the invasion of *Carpobrotus edulis* in Uruguay based on citizen science records. Biological Invasions 26: 935–942. https://doi.org/10.1007/s10530-023-03242-w

Guézou A, Trueman M, Buddenhagen CE, Chamorro S, Guerrero AM, Pozo P, Atkinson R (2010) An Extensive Alien Plant Inventory from the Inhabited Areas of Galapagos. PLOS ONE 5: e10276. https://doi.org/10.1371/journal.pone.0010276

Henrys PA, Mondain-Monval TO, Jarvis SG (2024) Adaptive sampling in ecology: Key challenges and future opportunities. Methods in Ecology and Evolution 15: 1483–1496. https://doi.org/10.1111/2041-210X.14393

Hochmair HH, Scheffrahn RH, Basille M, Boone M (2020) Evaluating the data quality of iNaturalist termite records. PLOS ONE 15: 1–19. https://doi.org/10.1371/journal.pone.0226534

Howard JP (2020) Phonetic Spelling Algorithm Implementations for R. Journal of Statistical Software 95: 1–21. https://doi.org/10.18637/jss.v095.i08

Hsieh TC, Ma KH, Chao A (2016) iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods in Ecology and Evolution 7: 1451–1456. https://doi.org/10.1111/2041-210X.12613

Inderjit, Pergl J, van Kleunen M, Hejda M, Babu CR, Majumdar S, Singh P, Singh SP, Salamma S, Rao BRP, Pyšek P (2018) Naturalized alien flora of the Indian states: Biogeographic patterns, taxonomic structure and drivers of species richness. Biological Invasions 20: 1625–1638. https://doi.org/10.1007/s10530-017-1622-y

Levenshtein V (1966) Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics, Doklady 10: 707–710.

Li D, Potgieter L, Aronson M, Axmanová I, Baiser B, Carboni M, Celesti-Grapow L, Knapp S, Kühn I, Lacerda De Matos AC, Lososová Z, Montaño-Centellas F, Pyšek P, Richardson D, Trotta L, Zenni R, Cilliers S, Clarkson B, Davis A, Dolan R, Dyderski M, Essl F, Gaoue O, Gui J, Géron C, Heringer G, Hui C, Khuroo A, Klotz S, Kotanen P, Kreft H, La Sorte F, Lembrechts J, Lenzner B, Lepczyk C, MacIvor S, Martínez-Garza C, Mori A, Nilon C, Pergl J, Siebert S, Tretyakova A, Tsang T, Uchida K, Van Kleunen M, Vilà M, Wang H-F, Weigelt P, Werner P, Williams N, Winter M, Cadotte M (2024) GUBIC: The global urban biological invasions compendium for plants. Ecological Solutions and Evidence 6: e70020. https://doi.org/10.32942/X20P8J

Louw GJ, Potgieter LJ, Richardson DM (2024) *Myoporum* (Scrophulariaceae): Introduction, naturalization, and invasion of an enigmatic tree genus in South Africa. South African Journal of Botany 168: 529–541. https://doi.org/10.1016/j.sajb.2024.03.022

McLean P, Wilson JRU, Gaertner M, Kritzinger-Klopper S, Richardson DM (2018) The distribution and status of alien plants in a small South African town. South African Journal of Botany 117: 71–78. https://doi.org/10.1016/j.sajb.2018.02.392

Mesaglio T (2024) A Guide to iNaturalist: An Australian Perspective. Atlas of Living Australia. https://doi.org/10.54102/ala.33279

Mesaglio T, Callaghan CT, Samonte F, Gorta SBZ, Cornwell WK (2023) Recognition and completeness: Two key metrics for judging the utility of citizen science data. Frontiers in Ecology and the Environment 21: 167–174. https://doi.org/10.1002/fee.2604

Mesaglio T, Kar F, Sauquet H, Cornwell WK (2025) infinitylists: A Shiny application and R package for rapid generation of place-based species checklists. Applications in Plant Sciences 13: e70012. https://doi.org/10.1002/aps3.70012

Milton SJ, Dean WRJ (2025) Village aliens: Self-seeding of indigenous and non-indigenous woody garden plants in a small desert village. Journal of Arid Environments 227: 105295. https://doi.org/10.1016/j.jaridenv.2024.105295

Nelufule T, Robertson MP, Wilson JRU, Faulkner KT (2023) An inventory of native-alien populations in South Africa. Scientific Data 10: 213. https://doi.org/10.1038/s41597-023-02119-w

Palit R, Downey PO, Duncan RP (2024) Factors influencing naturalisation success in horticultural species: A case study using planting records from the inception of a planned city. Biological Invasions. https://doi.org/10.1007/s10530-024-03281-x

Pickett STA, Cadenasso ML, Grove JM, Nilon CH, Pouyat RV, Zipperer WC, Costanza R (2001) Urban Ecological Systems: Linking Terrestrial Ecological, Physical, and Socioeconomic Components of Metropolitan Areas. Annual Review of Ecology and Systematics 32: 127. https://doi.org/10.1146/annurev.ecolsys.32.081501.114012

Potgieter LJ, Cadotte MW (2020) The application of selected invasion frameworks to urban ecosystems. NeoBiota 62: 365–386. https://doi.org/10.3897/neobiota.62.50661

Potgieter LJ, Douwes E, Gaertner M, Measey J, Paap T, Richardson DM (2020) Biological invasions in South Africa's urban ecosystems: Patterns, processes, impacts, and management. In: van Wilgen BW, Measey J, Richardson DM, Wilson JR, Zengeya TA (Eds) Biological Invasions in South Africa. Springer International Publishing, Cham, 275–309. https://doi.org/10.1007/978-3-030-32394-3_11

Potgieter LJ, ter Huurne MB, Richardson DM (2024a) Community science can inform invasive species management: *Melaleuca* (Myrtaceae) in South Africa. Ecological Solutions and Evidence 5: e12391. https://doi.org/10.1002/2688-8319.12391

Potgieter LJ, Cadotte MW, Roets F, Richardson DM (2024b) Monitoring urban biological invasions using citizen science: The polyphagous shot hole borer (*Euwallacea fornicatus*). Journal of Pest Science. https://doi.org/10.1007/s10340-024-01744-7

Potgieter LJ, Li D, Baiser B, Kühn I, Aronson MFJ, Carboni M, Celesti-Grapow L, de Matos ACL, Lososová Z, Montaño-Centellas FA, Pyšek P, Richardson DM, Tsang TPN, Zenni RD, Cadotte MW (2024c) Cities shape the diversity and spread of nonnative species. Annual Review of Ecology, Evolution, and Systematics 55.

Poynton RJ (2009) Tree Planting in Southern Africa: Vol 3. Other genera. Department of Agriculture, Forestry and Fisheries, Pretoria.

QGIS Development Team (2023) QGIS Geographic Information System. Open Source Geospatial Foundation Project. https://www.qgis.osgeo.org [March 9, 2023]

R Core Team (2025) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rega-Brodsky CC, Aronson MFJ, Piana MR, Carpenter ES, Hahs AK, Herrera-Montes A, Knapp S, Kotze DJ, Lepczyk CA, Moretti M, Salisbury AB, Williams NSG, Jung K, Katti M, MacGregor-Fors I, MacIvor JS, La Sorte FA, Sheel V, Threfall CG, Nilon CH (2022) Urban biodiversity: State of the science and future directions. Urban Ecosystems 25: 1083–1096. https://doi.org/10.1007/s11252-022-01207-w

Richardson DM, Potgieter LJ (2024) A living inventory of planted trees in South Africa derived from iNaturalist. South African Journal of Botany 173: 365–379. https://doi.org/10.1016/j.sajb.2024.08.006

Richardson DM, Pyšek P, Rejmánek M, Barbour MG, Panetta FD, West CJ (2000) Naturalization and invasion of alien plants: Concepts and definitions. Diversity & Distributions 6: 93–107. https://doi.org/10.1046/j.1472-4642.2000.00083.x

Richardson DM, Foxcroft LC, Latombe G, Le Maitre DC, Rouget M, Wilson JR (2020) The Biogeography of South African terrestrial plant invasions. In: Van Wilgen BW, Measey J, Richardson DM, Wilson JR, Zengeya TA (Eds) Biological Invasions in South Africa. Springer International Publishing, Cham, 67–96. https://doi.org/10.1007/978-3-030-32394-3_3

Roy HE, Pauchard A, Stoett PJ, Renard Truong T, Meyerson LA, Bacher S, Galil BS, Hulme PE, Ikeda T, Kavileveettil S, McGeoch MA, Nuñez MA, Ordonez A, Rahlao SJ, Schwindt E, Seebens H, Sheppard AW, Vandvik V, Aleksanyan A, Ansong M, August T, Blanchard R, Brugnoli E, Bukombe JK, Bwalya B, Byun C, Camacho-Cervantes M, Cassey P, Castillo ML, Courchamp F, Dehnen-Schmutz K, Zenni RD, Egawa C, Essl F, Fayvush G, Fernandez RD, Fernandez M, Foxcroft LC, Genovesi P, Groom QJ, González AI, Helm A, Herrera I, Hiremath AJ, Howard PL, Hui C, Ikegami M, Keskin E, Koyama A, Ksenofontov S, Lenzner B, Lipinskaya T, Lockwood JL, Mangwa DC, Martinou AF, McDermott SM, Morales CL, Müllerová J, Mungi NA, Munishi

LK, Ojaveer H, Pagad SN, Pallewatta NPKTS, Peacock LR, Per E, Pergl J, Preda C, Pyšek P, Rai RK, Ricciardi A, Richardson DM, Riley S, Rono BJ, Ryan-Colton E, Saeedi H, Shrestha BB, Simberloff D, Tawake A, Tricarico E, Vanderhoeven S, Vicente J, Vilà M, Wanzala W, Werenkraut V, Weyl OLF, Wilson JRU, Xavier RO, Ziller SR (2024) Curbing the major and growing threats from invasive alien species is urgent and achievable. Nature Ecology & Evolution 8: 1216–1223. https://doi.org/10.1038/s41559-024-02412-w

Saffer A, Worm T, Takeuchi Y, Meentemeyer R (2024) GIATAR: A Spatio-temporal Dataset of Global Invasive and Alien Species and their Traits. Scientific Data 11: 991. https://doi.org/10.1038/s41597-024-03824-w

Salomon Cavin J, Kull CA (2017) Invasion ecology goes to town: From disdain to sympathy. Biological Invasions 19: 3471–3487. https://doi.org/10.1007/s10530-017-1588-9

Seebens H, Kaplan E (2022) DASCO: A workflow to downscale alien species checklists using occurrence records and to re-allocate species distributions across realms. NeoBiota 74: 75–91. https://doi.org/10.3897/neobiota.74.81082

Seebens H, Meyerson LA, Richardson DM, Akite P, Alexander JM, Bailey SA, Biancolini D, Blackburn TM, Boehmer HJ, Bortolus A, Cadotte MW, Capinha C, Courchamp F, Crouch JA, Daehler CC, Essl F, Foxcroft LC, Fridley JD, Fuentes N, Gaertner M, Garcia-Berthou E, Garcia-Diaz P, Haider S, Hughes KA, Hui C, Kaplan E, Keshin E, Lenzner B, Liebhold AM, Liu C, Marchante E, Marchante H, Marticorena A, Moreno RA, Nentwig W, Niamir A, Novoa A, Nuñez AL, Pauchard A, Ricciardi A, Russell J, Saeedi H, Sankaran KV, Schwindt E, Shackleton RT, Simberloff D, Strayer D, Thines M, Tricarico E, Villaseñor-Parada C, Vitule JRS, Wagner V, Werenkraut V, Wesche K, Willette DA, Zenni RD, Pyšek P (2025) Biological invasions: A global assessment of geographic distributions, long-term trends, and data gaps. Biological Reviews of the Cambridge Philosophical Society. https://doi.org/10.1111/brv.70058

Shen FY, Ding TS, Tsai JS (2023) Comparing avian species richness estimates from structured and semi-structured citizen science data. Scientific Reports 13: 1214. https://doi.org/10.1038/s41598-023-28064-7

Thompson MM, Moon K, Woods A, Rowley JJL, Poore AGB, Kingsford RT, Callaghan CT (2023) Citizen science participant motivations and behaviour: Implications for biodiversity data coverage. Biological Conservation 282: 110079. https://doi.org/10.1016/j.biocon.2023.110079

Van Kleunen M, Essl F, Pergl J, Brundu G, Carboni M, Dullinger S, Early R, González-Moreno P, Groom QJ, Hulme PE, Kueffer C, Kühn I, Máguas C, Maurel N, Novoa A, Parepa M, Pyšek P, Seebens H, Tanner R, Touza J, Verbrugge L, Weber E, Dawson W, Kreft H, Weigelt P, Winter M, Klonner G, Talluto MV, Dehnen-Schmutz K (2018) The changing role of ornamental horticulture in alien plant invasions. Biological Reviews of the Cambridge Philosophical Society 93: 1421–1437. https://doi.org/10.1111/brv.12402

Van Kleunen M, Pyšek P, Dawson W, Essl F, Kreft H, Pergl J, Weigelt P, Stein A, Dullinger S, König C, Lenzner B, Maurel N, Moser D, Seebens H, Kartesz J, Nishino M, Aleksanyan A, Ansong M, Antonova LA, Barcelona JF, Breckle SW, Brundu G, Cabezas FJ, Cárdenas D, Cárdenas-Toro J, Castaño N, Chacón E, Chatelain C, Conn B, de Sá Dechoum M, Dufour-Dror J-M, Ebel AL, Figueiredo E, Fragman-Sapir O, Fuentes N, Groom QJ, Henderson L, Inderjit, Jogan N, Krestov P, Kupriyanov A, Masciadri S, Meerman J, Morozova O, Nickrent D, Nowak A, Patzelt A, Pelser PB, Shu W, Thomas J, Uludag A, Velayos M, Verkhosina A, Villaseñor JL, Weber E, Wieringa JJ, Yazlık A, Zeddam A, Zykova E, Winter M (2019) The Global Naturalized Alien Flora (GloNAF) database. Ecology 100: e02542. https://doi.org/10.1002/ecy.2542

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the Tidyverse. Journal of Open Source Software 4: 1686. https://doi.org/10.21105/joss.01686

Wilson JRU, Faulkner KT, Rahlao SJ, Richardson DM, Zengeya TA, van Wilgen BW (2018) Indicators for monitoring biological invasions at a national level. Journal of Applied Ecology 55: 2612–2620. https://doi.org/10.1111/1365-2664.13251

Young BE, Lee MT, Frey M, Barnes K, Hopkins P (2021) Using Citizen Science Observations to Develop Managed Area Watch Lists. Natural Areas Journal 41: 307–314. https://doi.org/10.3375/21-8

Appendix 1

Defining Western Cape urban boundaries

The delineation of urban boundaries is challenging and context/region specific (Rega-Brodsky et al. 2022). Urban centres can be defined as areas with a high density of people and/or a large proportion of built infrastructure. These can include built-up areas and embedded urban green spaces (Pickett et al. 2001). In delineating the urban boundaries for the Western Cape, we had four objectives: 1) to apply a standardised protocol across all urban centres to enable comparison; 2) to include as much urban built-up area as possible; 3) to include as much intra-urban green spaces (e.g., parks, road verges, vacant lots) as possible; and 4) to exclude as much non-urban natural and agricultural areas as possible.

Boundaries of the Western Cape urban centres were delineated using the latest urban land cover dataset (generated in 2020) for the Western Cape from the South African National Land Cover database (available at: https://egis.environment.gov.za/sa_national_land_cover_datasets). Using the "built-up" feature class from this dataset, a 500 m buffer was drawn around all continuous built-up areas covering at least five hectares (5 000 m^2). This was dissolved, i.e. all overlapping buffered areas were merged, before a negative 500 m buffer was applied. This process effectively removed internal boundaries between adjacent built-up areas in close proximity, incorporating intra-urban non-built-up areas such as parks and road verges, and smoothing the outer boundaries while preserving the general shape of the urban centre in alignment with the original "built-up" feature class outline.

Any erroneous "towns" including power infrastructure, military bases, and airports were removed. Though these could be considered urban centres they are unlikely to have good coverage of iNaturalist records (due to restricted access) and were not considered in this study. The remaining urban centres were ranked by size, and the 100 largest were retained. Urban centres ranged from ~ 821 km^2 (Cape Town) to 1.7 km^2 (Bot River) in extent. Smaller areas below our cut-off would have been difficult to differentiate from large agricultural infrastructure and were unlikely to have sufficient iNaturalist coverage. Fig. 1 shows the urban centres included in this study across the Western Cape.

The shapefiles of the urban centres were grouped according to local districts and uploaded to iNaturalist as "places". Only the City of Cape Town Metropolitan Municipality was handled differently to accommodate restrictions on place uploads imposed by iNaturalist. Because of the large number of records in this area, it was separated into smaller areas before uploading. All manipulations were performed in QGIS version 3.28.3 (QGIS Development Team 2023). The project is available at: (https://www.inaturalist.org/projects/vascular-plants-of-western-cape-urban-areas, accessed 23 October 2024).

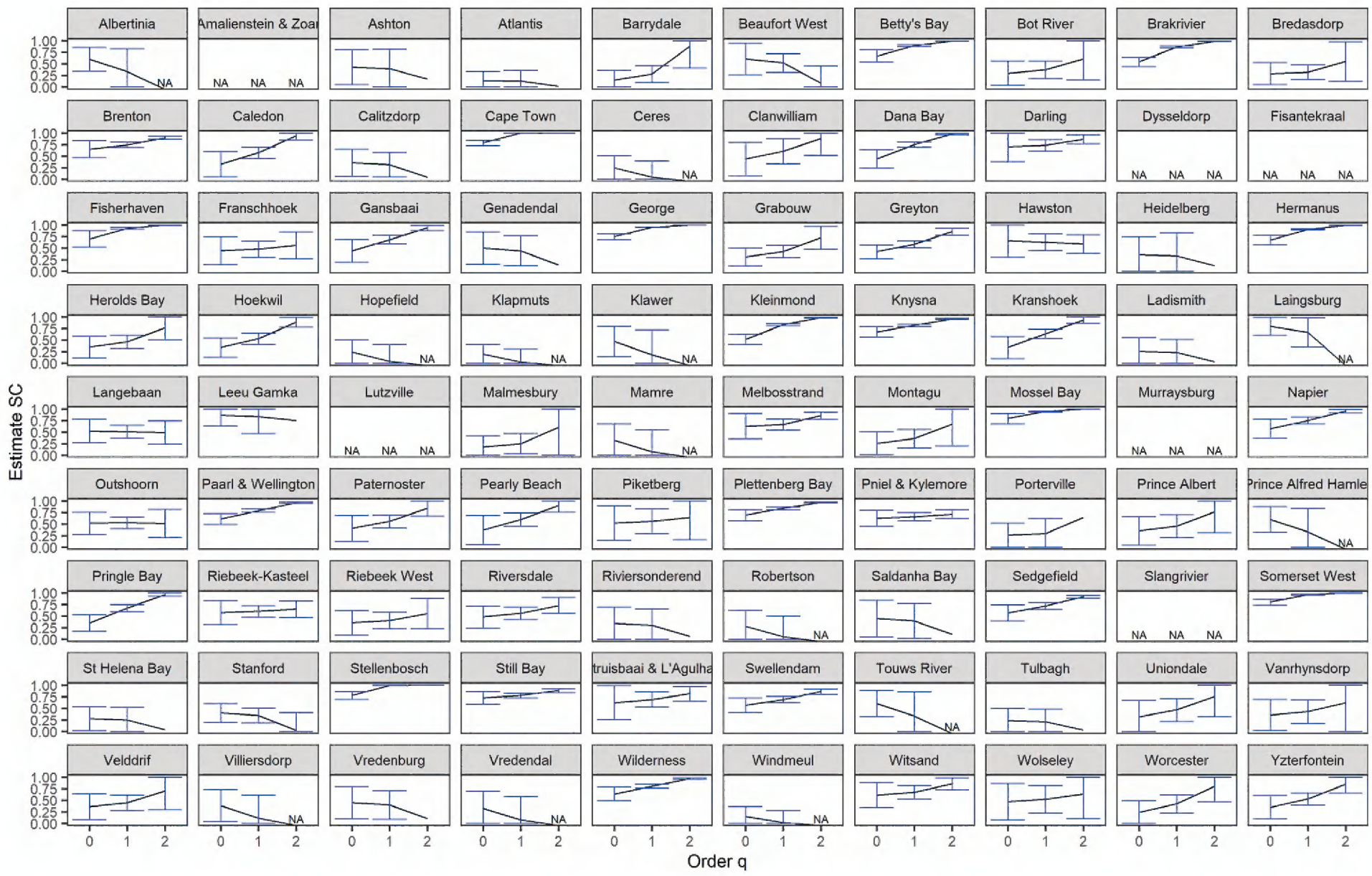
Appendix 2

Table A1. Plant taxa that could not be identified to species level in iNaturalist records for the Western Cape, South Africa. The potential/suggested species names are the taxa most often suggested (listed in the order of the most commonly suggested) by iNaturalist users (and known to be present in the region), ranked according to their frequency in the dataset (https://www.inaturalist.org/projects/naturalising-alien-plants-of-western-cape-urban-areas [accessed: 3 November 2024]).

Taxon	Rank	Reason	Potential/suggested species names
Alocasia	Genus	Unclear distinction between various species	*A. macrorrhizos*, *A. brisbanensis*, *A. Odora*
Amaranthus	Genus	Unclear distinction between many species (there are several native species, but most taxa in urban centres are most likely non-native)	*A. deflexus*, *A. curentus*, *A. hybridus*, *A. viridis*, *A. blitum*
Avena	Genus	Species impossible to distinguish from iNaturalist images (details of spikelet morphology)	*A. fatua*, *A. barbata*, *A. sativa*, *A. Sterilis*
Bambusoideae	Subfamily	Many taxa; difficult to distinguish without flowers	*Phyllostachys aurea*, *Bambusa vulgaris*, *B. balcooa*
*Brugmansi*a	Genus	Difficult to distinguish species from typical iNaturalist images	*B. suaveolens*, *B. arborea*, *B. insignis*, *B.* × *candida*
Eucalyptus	Genus	Species difficult/impossible to identify at juvenile stage or without flowers/fruit	*E. conferruminata*, *E. camaldulensis*, *E. cladocalyx*, *E. botryoides*, *E. gomphocephala*, *E. cinerea*, *E. grandis*, *E. globulus*, *E. polyanthemos*, *E. microcorys*, *E. melliodora*
Fraxinus	Genus	Unclear distinction between species	*F. pennsylvanica*, *F. excelsior*, *F. americana*, *F. angustifolia*
*Gamochaet*a	Genus	Unclear species distinctions	*G. americana*, *G. pensylvanica*
Hedera	Genus	Unclear distinction between *H. helix* and *H. canariensis* without juvenile foliage	*H. helix*, *H. canariensis*
Narcissus	Genus	Possible hybridisation and several cultivars	*N. tazetta*, *N. papyraceus*
Neltuma	Genus	Difficult to distinguish	*N. glandulosa*, *N. juliflora*, *N. odorata*, *N. velutina*
Oenothera	Section	Indistinguishable species and hybridisation	*O. drummondii*, *O. stricta*, *O. indecora*, *O. biennis*, *O. laciniata*, *O. affinis*, etc.
Pinus	Genus	Difficult to identify at juvenile stage and from image quality typically available on iNaturalist	*P. radiata*, *P. pinaster*, *P. pinea*, *P. halepensis*, *P. canariensis*
Sagina	Genus	Unclear species distinctions	*S. apetala*, *S. micropetala*, *S. Procumbens*
Salsola	Genus	Difficult to distinguish	*S. kali*, *S. tragus*, *S. australis*
Sambucus	Genus	Two species difficult to distinguish	*S. nigra*, *S. canadensis*
Vitis	Genus	Most likely *V. vinifera* but other difficult to distinguish species possibly present	*V. vinifera*, (*V. labrusca*)

Appendix 3

Completeness profiles for all urban centres with more than two records. See Fig. 4 for selected examples.



Supplementary material 1

The most abundant families in the inventory

Authors: Christiaan P. Gildenhuys, Luke J. Potgieter, Cang Hui, David M. Richardson

Data type: docx

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/neobiota.104.155832.suppl1